This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07K 14/00

A2

(11) International Publication Number: WO 99/09054

(43) International Publication Date: 25 February 1999 (25.02.99)

(21) International Application Number: PCT/BE98/00124

(22) International Filing Date: 20 August 1998 (20.08.98)

(30) Priority Data:
9700692 20 August 1997 (20.08.97) BE

(71) Applicants *(for all designated States except US)*: UNIVERSITE CATHOLIQUE DE LOUVAIN [BE/BE]; Halles Universitaires, Place de l'Université 1, B-1348 Louvain-la-Neuve (BE). UNIVERSITE DE MONS-HAINAUT [BE/BE]; Place du Parc 20, B-7000 Mons (BE).

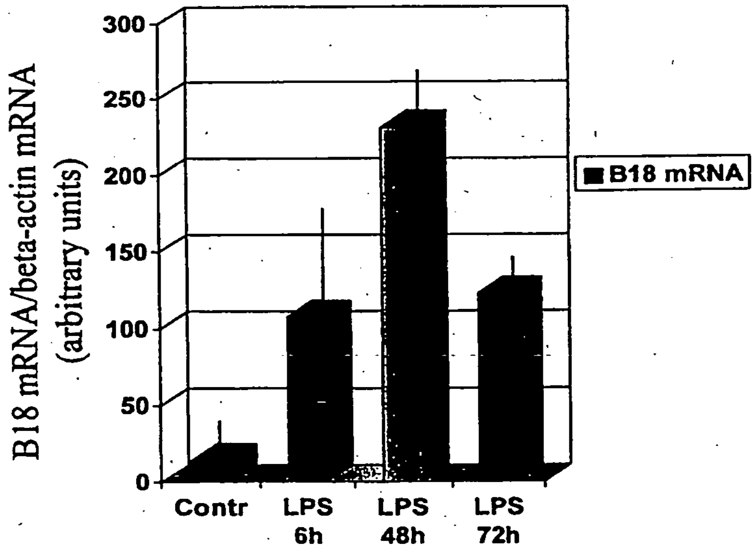
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KNOOPS, Bernard [BE/BE]; Rue Chapelle Notre-Dane 3/1, B-1341 Ceroux-Mousty (BE). HERMANS, Cedric [BE/BE]; Avenue des Glycines 42, B-1030 Brussels (BE). BERNARD, Alfred [BE/BE]; Avenue de la Chapelle 6, B-1200 Brussels (BE). WATTIEZ, Ruddy [BE/BE]; Chemin du Sauvelon 17, B-7022 Hyon (BE). FALMAGNE, Paul [BE/BE]; Rue du Point du Jour 8, B-7022 Mesvin (BE).

(74) Agents: VAN MALDEREN, Eric et al.; Office Van Malderen, Place Reine Fabiola 6/1, B-1083 Brussels (BE).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PEROXISOME-ASSOCIATED POLYPEPTIDE, NUCLEOTIDE SEQUENCE ENCODING SAID POLYPEPTIDE AND THEIR USES IN THE DIAGNOSIS AND/OR THE TREATMENT OF LUNG INJURIES AND DISEASES, AND OF OXIDATIVE STRESS-RELATED DISORDERS



(57) Abstract

The present invention is related to an isolated and purified polypeptide which amino acid sequence presents more than 70 % with the sequence SEQ ID NO 1. The present invention is also related to the nucleotide sequence encoding said amino acid sequence, the inhibitor directed against said sequences and their use in the diagnosis, treatment and/or prevention of lung injuries or diseases and oxidative stress-related disorders.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

PEROXISOME-ASSOCIATED POLYPEPTIDE, NUCLEOTIDE SEQUENCE ENCODING SAID POLYPEPTIDE AND THEIR USES IN THE DIAGNOSIS AND/OR THE TREATMENT OF LUNG INJURIES AND DISEASES, AND OF OXIDATIVE STRESS-RELATED DISORDERS

Field of the invention

The present invention is related to a new peroxisome-associated polypeptide, the nucleotide sequence encoding said polypeptide and portions thereof as well as their uses in the diagnosis of several diseases, especially the diagnosis and/or the treatment of lung injuries and diseases, and of oxidative stress-related disorders.

Background of the invention

The peroxisomes are organelles nearly ubiquitous in eukaryotic cells. They contain enzymes essential for various catabolic and anabolic pathways. Some of these enzymes are expressed constitutively while others can be induced under appropriate conditions. Peroxisomes carry out a variety of essential reactions such as peroxisomal oxidation and respiration, fatty acid beta-oxidation, cholesterol and dolichol metabolism, ether-phospholipid synthesis, and glyoxylate and pipecolic acid metabolism.

The peroxisomal respiratory pathway is based upon the formation of hydrogen peroxide by a collection of oxidases and the decomposition of the H_2O_2 by catalase. These reactions are responsible for 20% of oxygen consumption in liver, and several oxidases have been identified in peroxisomes. Ethanol elimination via catalase in peroxisomes may be significant in addition to the oxidation via cytosolic alcohol dehydrogenase.

The peroxisomal beta-oxidation system catalyses the beta-oxidative chain shortening of a specific set of compounds which can not be handled by mitochondria : very long chain fatty acids, di- and trihydroxycholestanoic acids, pristanic acid, long chain dicarboxylic acids, several prostaglandins, several leukotrienes, 12- and 15-hydroxyeicosatetraeonic acid, and several mono- and polyunsaturated fatty acids, which are of direct diagnostic relevance for some peroxisomal disorders.

Peroxisomes play also a major role in the synthesis of cholesterol and other isoprenoids. Fibroblasts from patients affected by disorders of peroxisome biogenesis show low capacity to synthesise cholesterol.

Two enzyme activities responsible for introduction of the characteristic ether linkage in ether-linked phospholipids (dihydroacetonephosphate acyltransferase (DHAPAT) and alkyldihydroxyacetonephosphate synthase (alkyl-DHAP synthase)) are localised in peroxisomes. These enzymes are not yet cloned. As demonstrated by the identification of patients with deficiency of either DHAPAT or alkyl-DHAP synthase with severe clinical abnormalities, ether-phospholipids are of major importance in humans.

Peroxisomes are able to detoxify glyoxylate via alanine/glyoxylate aminotransferase. The deficiency of this cloned enzyme causes hyperoxaluria type I.
L-pipecolate is a minor metabolite of L-lysine and is catabolised by the L-pipecolate oxidase localised in peroxisomes. The enzyme is deficient in cerebro-hepato-renal (Zellweger) syndrome.

In human, the importance of peroxisomes was emphasised by a number of inherited diseases involving either a defect in the biogenesis of peroxisomes or a deficiency of one (or more) peroxisomal enzymes. So far, 12 different peroxisomal disorders have been described and most of them are lethal.

A wide variety of chemicals have been showed to produce peroxisome proliferation and induction of peroxisomal and microsomal fatty acids-oxidising enzymes activities in rats and mice. Several peroxisomes proliferators have been shown to increase the incidence of liver tumours in these species. Proposed mechanisms of liver tumour formation by peroxisomes proliferators include induction of sustained oxidative stress.

Therefore, newly identified molecules associated with peroxisomes could be used for the development of diagnostic tools and possibly for the improvement of several therapeutical applications of various diseases associated with peroxisomal disorders. In addition, it is useful to identify the molecules present in specific organs like the lung and which may be used as specific markers of inflammatory diseases as well as lung injuries or diseases.

Summary of the invention

The Inventors have isolated and purified a new sequence of a low molecular weight human broncho-alveolar polypeptide. Said mammal, preferably human, protein or polypeptide (hereafter identified as B18hum protein) has been sequenced and its corresponding genomic DNA (SEQ ID NO 8) and cDNA (SEQ ID NO 1) have been identified. Similarly, the corresponding nucleotide and amino acid sequence from a rat (SEQ ID NO 3 and 4) and from a mouse (SEQ ID NO 5 and 6) have been obtained.

Said sequences present several homologies with other peroxisomal proteins of yeast and comprise a carboxy-terminal tripeptide SQL which is necessary for the specific targeting and translocation of several proteins into the peroxisome.

Therefore, the present invention is related to a new isolated and purified polypeptide sequence having a amino acid sequence which presents more than 70% homology, advantageously more than 85% homology, more preferably more than 95% homology, with the amino acid sequence SEQ ID NO 2., Said amino acid sequence is advantageously obtained from a mammal, preferably from a rat, a mouse or a human.

The present invention is also related to the isolated and purified polypeptide sequence corresponding to the amino acid sequence SEQ ID NO 2 or a portion thereof, preferably an immunoreactive portion (putative immunogenic domain or T or B cell epitopes).

Said portions are advantageously comprised between :

- Glutamic acid position 13 - Glutamic acid position 27
- Alanine position 26 - Leucine position 36

- Alanine position 42 - Glutamic acid position 57
- Glutamic acid position 57 - Valine position 69
- Valine position 80 - Leucine position 97
- Arginine position 95 - Leucine position 112
- Serine position 118 - Serine position 129
- Valine position 137 - Threonine position 150

Preferably, said portion has more than 10, 20, 30, 50 or 70 amino acids. Specific portions of the amino acid sequence SEQ ID NO 2 are also portions of more than 70 amino acids which present at least 80% of the proteinic activity (see example 5) of the complete SEQ ID NO 2 sequence. Therefore, the amino acid sequence according to the invention can be partially deleted while maintaining its activity, preferably its anti-oxidative activity, which will be described hereafter.

According to the invention, the amino acid sequence SEQ ID NO 2 presents a pI of 7.16 and a molecular weight of 17047 Dalton as hereafter defined by bidimensional electrophoresis.

The present invention is also related to the nucleotide sequence encoding the amino acid sequence according to the invention and its regulatory sequences upstream said coding sequence. A nucleotide sequence encoding the polypeptide according to the invention is a genomic DNA (see SEQ ID NO 10), a cDNA (see SEQ ID NO 1) or a mRNA, possibly comprising said upstream regulatory sequence. Advantageously, said nucleotide sequence presents more than 70%, advantageously more than 85%, more preferably more than 95% homology with SEQ ID NO 1 or its complementary strand.

According to a preferred embodiment of the present invention, said nucleotide sequence corresponds to the nucleotide sequence SEQ ID NO 1, its complementary strand or a portion thereof.

"A portion of the nucleotide sequence SEQ ID NO 1" means any nucleotide sequence of more than 15 base pairs (such as a primer, a probe or an antisense nucleotide sequence) which allow the specific identification, reconstitution or blocking of the complete nucleotide sequence SEQ ID NO 1 (including its regulatory sequences upstream the coding sequence).

Said portions allow the specific identification, reconstitution or blocking by specific hybridisation with the nucleotidic sequence SEQ ID NO 1, preferably under standard stringent conditions, with sequences like probes or primers possibly labelled with a compound (radioactive compound, enzyme, fluorescent marker, etc.), and can be used in a specific diagnostic or dosage method like probe hybridisation (see Sambrook et al., §§ 9.47-9.51 in *Molecular Cloning : A Laboratory Manual*, Cold Spring Harbor, Laboratory Press, Cold Spring Harbor, New York (1989)), genetic amplification (like PCR (US patent 4,683,195), LCR (Wu et al., *Genomics 4*, pp. 560-569), CPR (US patent 5,011,769)).

Exemplary stringent hybridisation conditions are as follows : hybridisation at 42 ^{O}C in 50% formamide 5x SSC, 20 mM sodium phosphate, pH 6.8 washing in 0.2x SSC at 55 ^{O}C. It is understood by those skilled in the art that variation of these conditions occur based on the length and GC nucleotide content of the sequence to be hybridised. Formulas standard in the art are appropriated for determining exact hybridisation conditions (see Sambrook et al.

Preferred examples of said nucleotide portions are as follows :

	Sequence	Position
5'-gccatcccagcagtggaggtgtttg-3'	(SEQ ID NO 11)	217-241
5'-ttgaacagctctgccaggttcacc-3'	(SEQ ID NO 12)	261-284
5'-tggaggtgtttgaaggggagccag-3'	(SEQ ID NO 13)	230-253
5'-caggttcaccttgttccctggctc-3'	(SEQ ID NO 14)	247-270
5'-gggtatgggactagctggcg-3'	(SEQ ID NO 15)	33-52
5'-ctggccaacattccaattgcag-3'	(SEQ ID NO 16)	747-768

and the sequences of respectively 601 (SEQ ID NO 8), 604 (SEQ ID NO 9) and 469 (SEQ ID NO 7) base pairs corresponding to specific mRNA alternative splicing of the B18 human nucleotide sequence as described in Figure 4 (the known genomic sequence incorporating several introns and exons is represented in the sequence SEQ ID NO 10).

Said sequences may be used for a genetic amplification or a probe hybridisation as above-described.

The present invention is also related to a vector comprising the necessary elements for the injection, transfection or transduction of cells and having incorporated one or more of the nucleotide sequences according to the invention. The vector according to the invention is selected from the group consisting of viruses, plasmids, phagemides, cationic vesicles, liposomes or a mixture thereof. Said vector may comprise also one or more adjacent regulatory sequences (such as promoter(s), secretion and termination signal sequence(s)), advantageously operably linked to the nucleotide sequence according to the invention.

The present invention is also related to the cell transformed by said vector and expressing the polypeptide according to the invention.

The nucleotide sequence according to the invention can be also introduced in said cell by the formation of $CaPO_4$-nucleic acid precipitate, DEAE-dextran-nucleic acid complex or by electroporation.

Another aspect of the present invention is related to an inhibitor of the polypeptide according to the invention or the nucleotide sequence according to the invention (including the upstream sequences like promoter-operator regulatory sequence which may be inhibited by a cis- and/or transactivating repressor). Said inhibitor is advantageously an antibody or a fragment of said antibody such as an hypervariable portion of said antibody directed against the amino acid or nucleotide sequence of the polypeptide according to the invention. Other examples of inhibitors according to the invention are antisense nucleotide sequences which allow the blocking of the expression of the nucleotide sequence according to the invention.

Another aspect of the present invention is related to a diagnostic device (such as a diagnostic kit or a chromatographic column) comprising an element selected from the group consisting of the amino acid sequence of said polypeptide, its nucleotide sequence, and/or the inhibitor according to the invention or a fragment thereof as above-described. Said diagnostic device may comprise also necessary reactants and media for the diagnostic and/or dosage of the nucleotide and/or amino acid sequence of the polypeptide according to the invention, which are based upon the method selected from the group consisting of in situ hybridisation, hybridisation by labelled antibodies, especially RIA (Radio Immuno Assay) or ELISA (Enzymes Linked Immuno-Sorbent Assay) technologies, detection upon filter, upon solid support, in solution, in sandwich, upon gel, dot blot hybridisation, Northern blot hybridisation, Southern blot hybridisation, isotopic or non-isotopic labelling (by immunofluorescence or biotinilised probes), genetic amplification, (especially by PCR or LCR), double immunodiffusion technique, counter-electrophoresis technique, haemagglutination or a mixture thereof.

Another aspect of the present invention concerns a diagnosis method wherein a biological sample from the patient, such as cephalo-rachidian fluid, serum, blood, plasma, urine, broncho-alveolar lavage, stomach lavage, etc., is isolated from the patient, and is put in contact with the diagnostic device according to the invention for the diagnosis or the monitoring of an injury or a disease, preferably a lung injury or an oxidative stress-related disorder, affected by the presence of pro-oxidant agent or oxidative stress such as specific cardio-vascular diseases like arteriosclerosis, neurodegenerative disorders (Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis), apoptosis, inflammatory reactions, allergic reactions such as asthma, hay fever and eczema, high bone mass syndrome, osteopetrosis, osteoporosis-pseudoglioma syndrome, and Bardet-Biedl syndrome 1. Said diagnosis and monitoring upon one or more biological samples obtained from several tissues from the patient can be advantageously obtained by one or more of the methods above-described, which could be adapted according to the specific biological sample by the person skilled in the art.

Therefore, the product according to the invention could be used as a marker for the above-identified injuries, diseases or disorders in a broad spectrum of tissues as shown in the enclosed Figure 1.

A further aspect of the present invention is related to a pharmaceutical composition comprising a pharmaceutically acceptable carrier and an element selected from the group consisting of the nucleotide sequence, the amino acid sequence of the polypeptide according to the invention, the inhibitor directed against said sequences and/or one or more portions thereof.

A last aspect of the present invention is related to the use of the pharmaceutical composition according to the invention for the manufacture of a medicament in the treatment and/or the prevention of lung injuries and/or diseases or of oxidative stress-related disorders.

The present invention is also related to a prevention and/or treatment method of a patient, especially a human patient, preferably affected by lung injuries and/or diseases or by oxidative stress-related disorders, wherein a sufficient amount of the pharmaceutical composition according to the invention is administered to said patient in order to treat, avoid and/or reduce the symptoms of said injuries and/or diseases.

Other injuries and/or diseases which can be prevented and/or treated are injuries and/or diseases affected by the presence of pro-oxidant agents or oxidative stress, such as specific cardio-vascular diseases like arteriosclerosis, neurodegenerative disorders such as

Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, apoptosis and inflammatory reactions and some allergic reactions such as asthma, hay fever and eczema, high bone mass syndrome, osteopetrosis, osteoporosis-pseudoglioma syndrome, and Bardet-Biedl syndrome 1.

The pharmaceutically acceptable carrier according to the invention is any compatible non-toxic substance suitable for administering the composition according to the invention to a human patient. Pharmaceutically acceptable carriers according to the invention suitable for oral administration are the ones well known by the person skilled in the art, such as tablets, coated or non-coated pills, capsules, spray-gas, patches, gels, solutions or syrups. Pharmaceutically acceptable carriers vary according to the mode of administration (intravenous, intramuscular, subcutaneous, parenteral, etc.), and may comprise also adjuvants well known by the person skilled in the art to increase, reduce and/or regulate humoral, local and/or cellular response of the immune system.

The pharmaceutical composition according to the invention may be prepared by the methods, generally applied by the person skilled in the art in the preparation of various pharmaceutical compositions, wherein the percentage of the active compound/pharmaceutically acceptable carrier can vary within very large ranges, only limited by the tolerance of the patient to said pharmaceutical composition, and wherein the limits are particularly determined by the frequency of administration and the possible side-effects of the active compounds or its pharmaceutically acceptable carrier.

Another aspect of the invention is related to the use of the diagnostic device according to the invention for performing upon the patient or upon a biological fluid obtained from the patient, a diagnosis, a dosage and/or a monitoring of the above-mentioned injuries or diseases or oxidative stress-related disorders affecting the patient.

A further aspect of the present invention is related to a cell or a non-human animal, preferably a mammal such as a mouse or a rat, transformed by the vector according to the invention and overexpressing the polypeptide according to the invention, or a non-human animal, preferably a mammal such as a mouse or a rat, genetically modified by a partial or total deletion of its genomic sequence encoding the polypeptide according to the invention (knock-out non-human mammal) and obtained by methods well known by the person skilled in the art such as the one described by Kahn et al. (*Cell*, Vol. 92, pp. 593-596 (March 1998)).

Other examples of genetically modified non-human animals according to the invention may be a transgenic non-human animal comprising an inhibitor according to the invention, preferably an antisense nucleic acid sequence complementary to the nucleotide sequence according to the invention so placed as to be transcribed into antisense mRNA which is complementary to the nucleotide sequence according to the invention and which hybridises to said nucleotide sequence, thereby reducing or blocking its translation.

Further aspects of the present invention will be described in the enclosed non-limiting examples in reference to the following Figures.

Brief description of the drawings

Figure 1	represents a dot blot analysis of mRNA encoding the polypeptide according to the invention in various types of human tissues.
Figure 2	represents a Northern blot analysis of mRNA encoding the polypeptide according to the invention in a rat lung after administration of lipopolysaccharides (LPS) inducing an inflammatory reaction of the lung.
Figure 3	represents a Northern blot analysis of mRNA encoding the polypeptide according to the invention in a rat lung after intraperitoneal injection of pneumotoxicants.
Figure 4	is a schematic representation of the human genomic sequence, the complete cDNA sequence and the corresponding amino acid sequence.
Figure 5	represents respectively the alignment of the sequences of the human B18 polypeptide according to the invention with the corresponding rat and mouse sequences.

Example 1 : Homology between the B18 polypeptide according to the invention with other known nucleotide or amino acid sequences

The BLAST 2.0 software (gapped BLAST at the NCBI Internet site) was used for searching for homologies between human B18 (162 amino acids) and known polypeptides in databases (GenBank, SwissProt). Said search did not give perfect alignment with known peptides from different species (Table 1). Homologies of the human B18 cDNA (805 nucleotides) with GenBank, EMBL, DDBJ and PDB deposited nucleotide sequences (Table 2) and GenBank Expression Sequence TAGS (ESTs) were noted.

Table 1 : Homologies of the B18 proteins (162 amino acid) with other proteins

Name	NCBI ID	Identity (%) Homology (%)
Membrane protein (synechocystis sp.)	1652859	57/129(44%) 81/129(62%)
Peroxisomal-like protein (Aspergillus fumigatus)	2769700	56/176(31%) 90/176(50%)
Haein HI0572 hypothetical protein(Haemophilus influenzae)	1723174	53/146(36%) 80/146(54%)
PMP20 (Schizosaccharomyces pombe)	AJ002536	54/161(33%) 85/161(52%)
Peroxisomal membrane protein A (PMP 20)(Candida boidinii)	130360	59/170(34%) 89/170(51%)
Peroxisomal membrane protein B (PMP 20)(Candida boidinii)	130361	58/170(34%) 88/170(51%)
Putative peroxisomal protein PMP from yeast (Saccharomyces cerevisiae)	1709682	41/138(29%) 72/138(51%)
Alkylhydroperoxide reductase C22 protein (Escherichia coli)	P26427	36/126(28%) 58/126(45%)

Table 2

Name	Access NO	Identity
Human mRNA down-regulated in cells infected by adenovirus 5	U82616	259/263 (98%)
Human mRNA down-regulated in cells infected by adenovirus 5	U82615	300/321 (93%)

In the Table 2, an identity of 98% has been obtained with the alignment of 259 nucleotides of CDNA B18, which comprises in its totality 805 nucleotides, with 263 nucleotides of U82616 CDNA. A similar identity has been obtained with the U82615 sequence.

The sequence SEQ ID NO 1 comprising 805 nucleotides presents a homology with several EST sequences obtained from a human and from a mouse, having the following references :

Human :

AA130751, N42215, W38597, N91311, N68467, AA187737, N68916, W00593, R88950, AA181884, H20154, H66666

Mouse :

AA220019, AA123351, AA087129, AA255021, AA249897, W71344

Example 2 : Tissue detection

A human RNA master Blot (Clontech) containing 100-500 ng of poly-A + human RNA in each dot (normalised to the mRNA expression levels of eight different housekeeping genes) was hybridised with a 554 bp-long B18 probe labelled with ^{32}P, and quantified using Phosphorimaging Technology. As shown in Figure 1, B18 mRNA is present in all tissues examined but predominantly in trachea, lung, kidney, thyroid gland, stomach, colon, heart and some regions of the brain. Highest expression has been noted in the thyroid tissue. This presence is probably correlated with the possible antioxidant activity of the B18 polypeptide according to the invention.

Example 3 : Inflammatory reaction

Figure 2 represents a Northern blot analysis of rat lung mRNA after 6, 48 and 72 hours after lipopolysaccharides (LPS) instillation inducing an inflammatory reaction in the lung.

A Northern blot containing 15 μg of total RNA in each lane was hybridised with a 225 bp-long rat B18 probe, stripped and reprobed with a 572 bp-long rat β-actin probe, both labelled with ^{32}P. Northern blot was quantified using Phosphorimaging Technology and the B18 mRNA data were normalised to β-actin mRNA level.

Example 4 : **Pneumotoxic reaction**

Figure 3 represents a Northern blot analysis of rat lung mRNA after intraperitoneal injection of pneumotoxicants (4-ipomeanol,1-(3-fyryl)-4-hydroxypentanone (IPO), methylcyclopentadienyl manganese tricarbonyl (MMT) or alpha naphtylthiourea (ANTU)). These agents are known to induce in the lung acute lesions of Clara (IPO) and alveolar cells (MMT) as well as increasing the permeability of the alveolar/blood barrier (ANTU). A Northern blot containing 15 μg of total RNA in each lane was hybridised with a 225 bp-long rat B18 probe, stripped and reprobed with a 572 bp-long β-actin probe both labelled with ^{32}P. The Northern blot was quantified using Phosphorimaging Technology and rat B18 mRNA data were normalised to β-actin mRNA level.

Example 5 : **Proteinic activity of the B18 polypeptide**

An amino analysis of the complete human B18 amino acid sequence shows that said polypeptide presents specific portions showing an homology with other anti-oxidant enzymes (starting from a Leucine at position 36 until a Cysteine at position 47) and an other portion having an important homology with beta chains of ATP synthase (starting from a Glutamic acid at position 13 until a Glycine in position 38).

Furthermore, the B18 amino acid sequence according to the invention shows an important homology with an *Aspergillus fumigatus* allergen (34% identity and 60% homology by using clustal V sequence alignment), especially upon the portion of said B18 polypeptide having possible antioxidant properties. Therefore, it is possible that a peroxisomal protein (possibly homologous to B18 protein) is able to induce and to bind IgE from patients sensitised to *Aspergillus fumigatus* peroxisomal proteins after an induction of the patient immune system with *Aspergillus fumigatus* allergen. This mechanism can be compared to a reaction obtained with the manganese superoxide dismutase (MnSOD) wherein the human MnSOD is able to bind to IgE from patients sensitised to *Aspergillus fumigatus* MnSOD.

Furthermore, the Inventors have identified a portion of the B18 human polypeptide which presents an homology with a Cyclophilin-binding domain of *Candida boidinii* PMP20 (receptor of the immuno-suppressant drug cyclosporine A). Said possible Cyclophilin-binding domain is starting from the Threonine in position 150 until the Leucine in position 161.

Example 6 : **B18 human gene and mRNA alternative splicing**

As represented in the enclosed Figure 4, the Inventors have identified upon the genomic DNA (SEQ ID NO 10) 5 exons and 5 introns. By RT-PCR (using primers 5'-gggtatgggactagctggcg-3' and 5'-ctggccaacattccaattgcag-3') and according to the genomic sequence, 4 different cDNAs corresponding to the transcription of the said genomic DNA have been identified in human lung and in human brain. A first cDNA of 736 bp corresponds to the cDNA encoding the complete amino acid sequence of the B18 protein according to the invention. However, 3 other cDNAs of 601, 604 and 469 bp were also identified, and comprise specific splicings of one or more exons.

Therefore, another aspect of the present invention is related to said specific portions of the complete genomic or CDNA nucleotide sequence according to the invention or to specific portions of the complete amino acid sequence of the B18 protein according to the invention, which could be used also as specific markers of the B18 activity, preferably the anti-oxidative activity.

Example 7 : Knock-out mouse

Exons of a mouse genomic sequence encoding the B18 polypeptide according to the invention have been deleted by homologous recombination. Said homologous recombination has been obtained with a genetic sequence comprising a neomycin resistant gene. The targeting vector with said gene and a thymidine kinase (in order to eliminate non-homologous recombinants with ganciclovir) has been prepared. Said recombination was used for the deletion of one or more exons of the B18 polypeptide. After electroporation of ES cells with the targeting vector, positive clones having incorporated homologous recombination were identified by Southern blot with labelled probes. Aggregation of said positive clones with a morula from a Swiss pseudo-pregnant mouse produces several chimeric mice which survive after birth. Several homozygote mice are obtained by cross-breeding and are used as a model for the above-mentioned diseases.

Similar experiments may be done with another mammal whose B18 sequence is known (the B18 sequence of a mouse and a rat and their alignment with the human sequence is shown in the enclosed Figure 5).

Example 8 : **Chromosome localisation of human B18 gene**

Radiation hybrid clones (GeneBridge 4 Radiation Hybrid Panel, Research Genetics) were used for performing chromosome localisation by PCR with two different pairs of primers (5'-caggttcaccttgttccctggctc-3' (SEQ ID NO 14), 5'-atgttatgcaaccctttgcgacac-3' (SEQ ID NO 17) and 5'-gtgtttgaaggggagccagggaac-3' (SEQ ID NO 18), 5'-agagacagggtttcaccatcttgg-3' (SEQ ID NO 19)).

The Inventors have located B18 genomic sequence on human chromosome 11q13. B18 gene has been located 7.15-6.1 cR from marker D11S913 between markers D11S1963 and D11S4407 (Genome Database internet site).

Unknown genes linked to different disorders have been localised in the same region of chromosome 11. Therefore, B18 gene is possibly associated with these disorders:

- atopy (atopic hypersensitivity: asthma, hay fever and eczema; MIM n°147050 at OMIM of NCBI internet site),
- high bone mass syndrome(MIM n°601884),
- osteopetrosis (MIM n°259700),
- osteoporosis-pseudoglioma syndrome (MIM n°259770) and
- Bardet-Biedl syndrome 1 (MIM n°209901).

CLAIMS

1. Amino acid sequence having more than 70% homology with the sequence SEQ ID NO 2.

2. Amino acid sequence according to claim 1, having more than 85% homology with the sequence SEQ ID NO 2.

3. Amino acid sequence according to claim 1 or 2, having more than 95% homology with the sequence SEQ ID NO 2.

4. Amino acid sequence according to any one of the preceding claims, corresponding to SEQ ID NO 2 or an immunoreactive portion thereof.

5. Nucleotide sequence encoding the amino acid sequence according to any one of the preceding claims and presenting more than 70% homology with SEQ ID NO 1 or its complementary strand.

6. Nucleotide sequence according to claim 5, having more than 85% homology with the sequence SEQ ID NO 1 or its complementary strand.

7. Nucleotide sequence according to claim 5 more than 95% homology with the sequence SEQ ID NO 1 or its complementary strand.

8. Nucleotide sequence according to any one of the claims 5 to 7, corresponding to the sequence SEQ ID NO 1, its complementary strand or a portion thereof specific for SEQ ID NO 1 and comprising more than 15 base pairs.

9. Vector comprising the nucleotide sequence according to any one of the claims 5 to 8.

10. Inhibitor directed against the amino acid or nucleotide sequence according to any one of the claims 1 to 8.

11. Inhibitor according to claim 10, being an antibody, preferably a monoclonal antibody, or a portion of said antibody.

12. Diagnostic device comprising an element selected from the group consisting of the amino acid sequence according to any one of the claims 1 to 4, the nucleotide sequence according to any one of the claims 5 to 8, the inhibitor according to claim 10 or 11, their portions or a mixture thereof.

13. Method for the in vitro detection of lung injuries and diseases or oxidative stress-related diseases and disorders, especially inflammatory diseases, comprising the steps of :

- isolating a sample from a body fluid of a patient, preferably a human patient,
- possibly inhibiting the contaminants present in said sample,
- put in contact said sample with an element selected from the group consisting of the amino acid sequence according to any one of the claims 1 to 4, the nucleotide sequence according to any one of the claims 5 to 8, the inhibitor according to claim 10 or 11, their portions or a mixture thereof, and
- detecting a reaction of a molecule present in said sample with said element.

14. Pharmaceutical composition comprising a pharmaceutically acceptable carrier and an element selected from the group consisting of the amino acid sequence according to any one of the claims 1 to 4, the nucleotide sequence according to any one of the claims 5 to 8, the inhibitor according to claim 10 or 11, their portions or a mixture thereof.

15. Use of the pharmaceutical composition according to claim 14 for the manufacture of a medicament for the prevention and/or the treatment of lung injuries or diseases, and of oxidative stress-related diseases or disorders, such as specific cardio-vascular diseases like arteriosclerosis, neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, apoptosis and inflammatory reactions, allergic reactions such as asthma, hay fever and eczema, high bone mass syndrome, osteopetrosis, osteoporosis-pseudoglioma syndrome, and Bardet-Biedl syndrome 1.

16. Cell transformed by the vector according to claim 9 or comprising a partial or total deletion of its nucleotide sequence according to any one of the claims 5 to 8.

17. Non-human animal, preferably a mammal, transformed by the vector according to claim 9 or comprising a partial or total deletion of its nucleotide sequence according to any one of the claims 5 to 8.

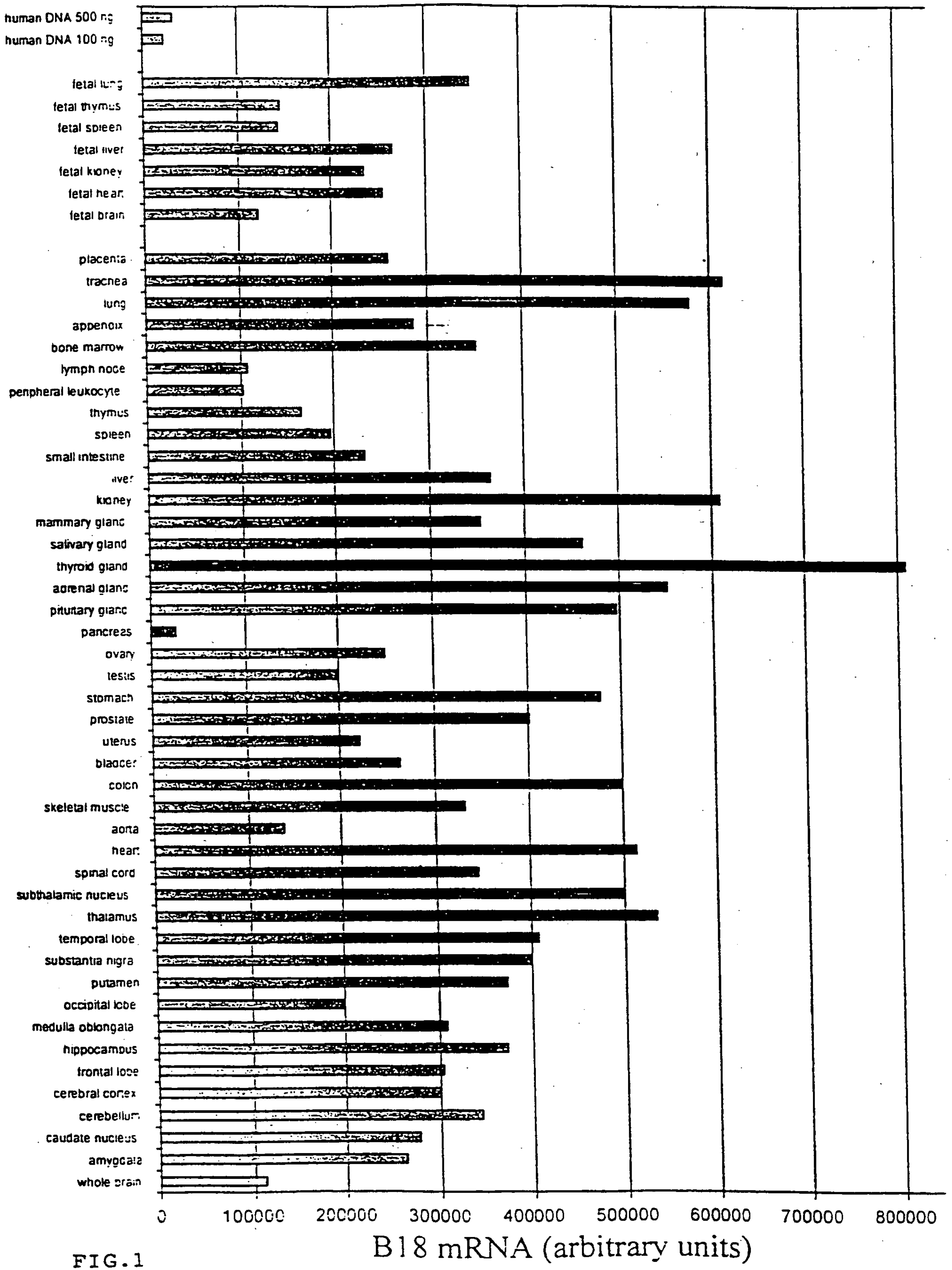

human DNA 500 ng
human DNA 100 ng
fetal lung
fetal thymus
fetal spleen
fetal liver
fetal kidney
fetal heart
fetal brain
placenta
trachea
lung
appendix
bone marrow
lymph node
peripheral leukocyte
thymus
spleen
small intestine
liver
kidney
mammary gland
salivary gland
thyroid gland
adrenal gland
pituitary gland
pancreas
ovary
testis
stomach
prostate
uterus
bladder
colon
skeletal muscle
aorta
heart
spinal cord
subthalamic nucleus
thalamus
temporal lobe
substantia nigra
putamen
occipital lobe
medulla oblongata
hippocampus
frontal lobe
cerebral cortex
cerebellum
caudate nucleus
amygdala
whole brain
0
100000
200000
300000
400000
500000
600000
700000
800000
B18 mRNA (arbitrary units)

FIG.1

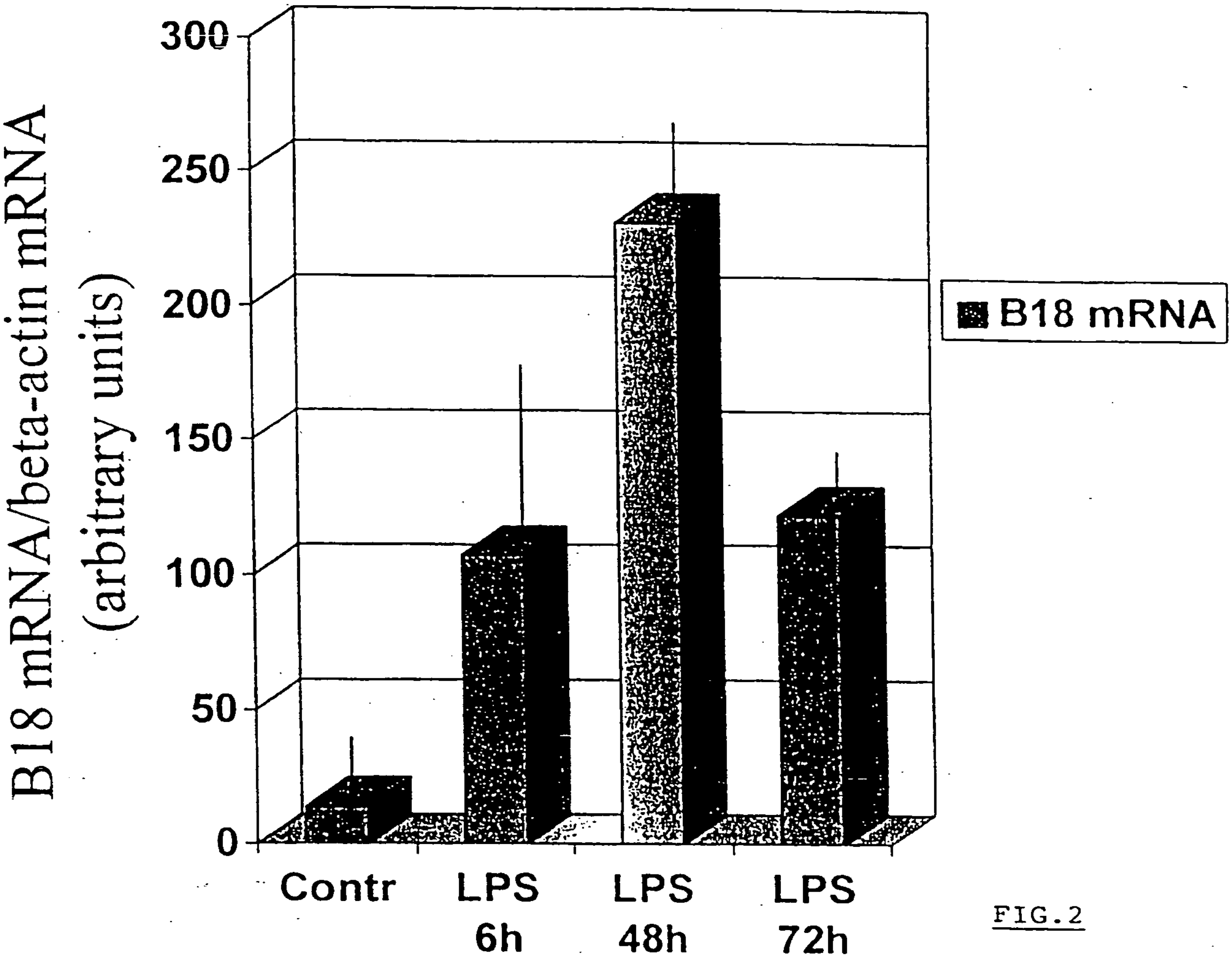

B18 mRNA/beta-actin mRNA
(arbitrary units)
300
250
200
150
100
50
0
Contr
LPS 6h
LPS 48h
LPS 72h
▩ B18 mRNA

FIG.2

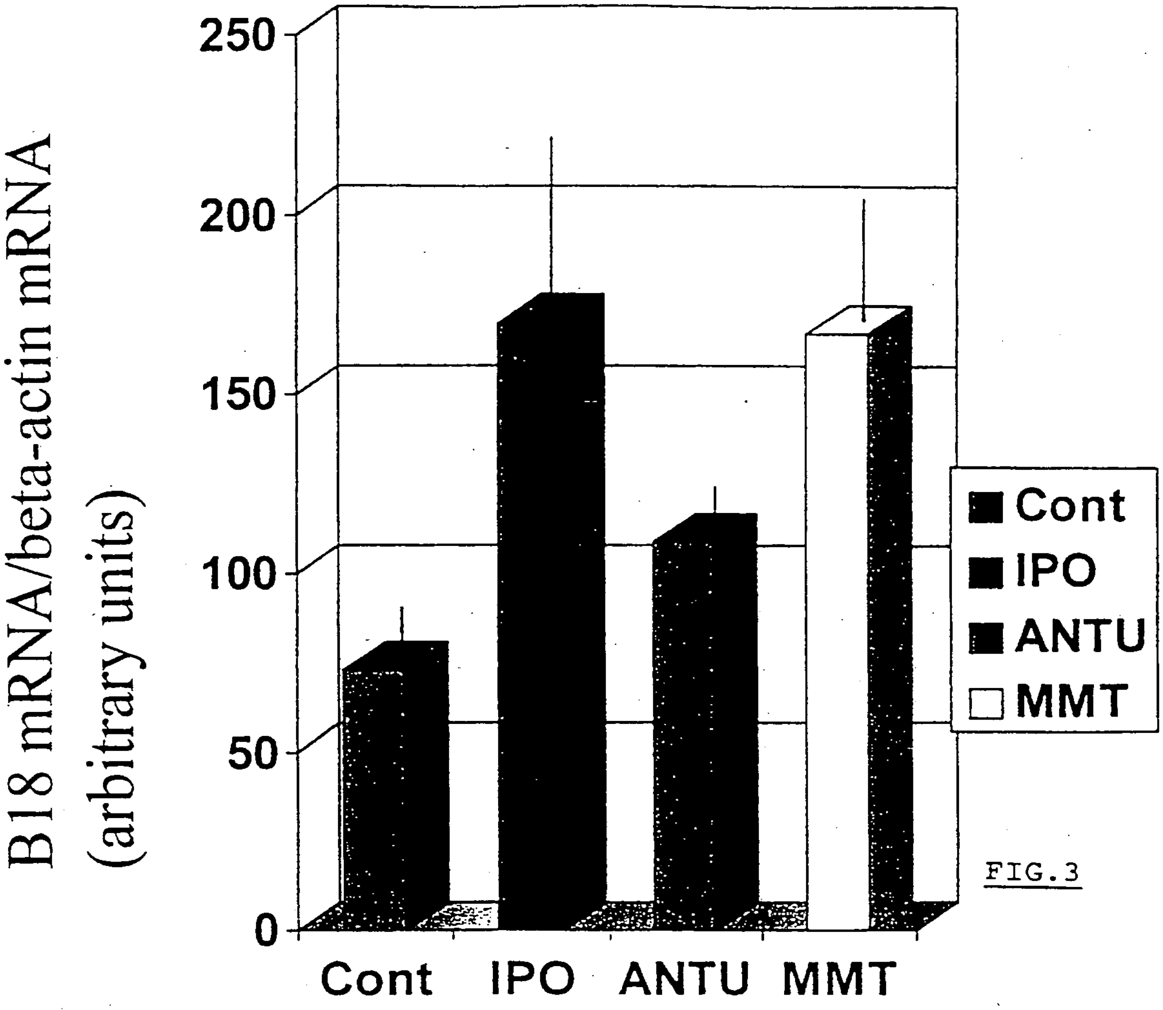

B18 mRNA/beta-actin mRNA
(arbitrary units)
250
200
150
100
50
0
Cont
IPO
ANTU
MMT
■ Cont
■ IPO
■ ANTU
□ MMT

FIG.3

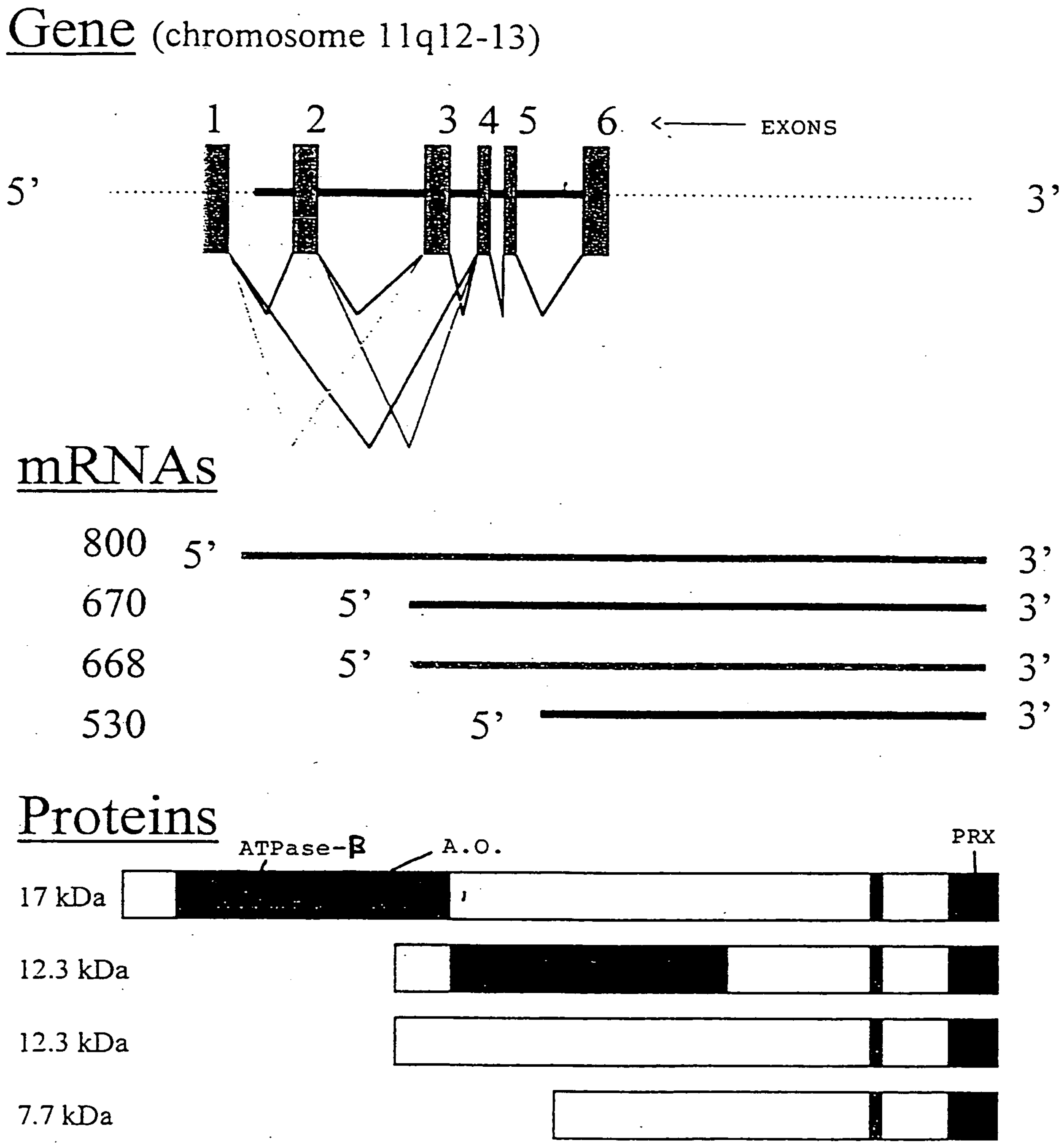

Gene (chromosome 11q12-13)
<— EXONS
mRNAs
Proteins
ATPase-β
A.O.
PRX
17 kDa
12.3 kDa
12.3 kDa
7.7 kDa

FIG.4

CLUSTAL V alignment of human and rat B18 amino acid sequences (Identity: 90%, Homology: 97.5%):

```
B18hum      MAPIKVGDAIPAVEVFEGEPGNKVNLAELFKGKKGVLFGVPGAFTPGCSK = SEQIDNO1
B18rat      MAPIKVGDTIPSVEVFEGEPGKKVNLAELFKDKKGVLFGVPGAFTPGCSK
            ********.**.*********.*********.******************

B18hum      THLPGFVEQAEALKAKGVQVVACLSVNDAFVTGEWGRAHKAEGKVRLLAD
B18rat      THLPGFVEQAGALKAKGAQVVACLSVNDVFVTAEWGRAHQAEGKVQLLAD   FIG.5a
            ********** ****** ********** ***.******.*****.****

B18hum      PTGAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQDGIVKALNVEPDGTGL
B18rat      PTGAFGKETDLLLDDSLVSLFGNRRLKRFSMVIDKGVVKALNVEPDGTGL
            *******************.************.. *.*************

B18hum      TCSLAPNIISQL
B18rat      TCSLAPNILSQL
            ********.***
```

CLUSTAL V alignment of human and mouse B18 amino acid sequences (Identity: 91%, Homology: 96%):

```
B18hum      MAPIKVGDAIPAVEVFEGEPGNKVNLAELFKGKKGVLFGVPGAFTPGCSK
B18mouse    MAPIKVGDAIPSVEVFEGEPGKKVNLAELFKGKKGVLFGVPGAFTPGCSK
            ***********.*********.****************************

B18hum      THLPGFVEQAEALKAKGVQVVACLSVNDAFVTGEWGRAHKAEGKVRLLAD
B18mouse    THLPGFVEQAGALKAKGAQVVACLSVNDVFVIEEWGRAHQAEGKVRLLAD
            ********** ****** ********** **  ******.**********

B18hum      PTGAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQDGIVKALNVEPDGTGL
B18mouse    PTGAFGKATDLLLDDSLVSLFGNRRLKRFSMVIDNGIVKALNVEPDGTGL
            ******* ***********.************...***************

B18hum      TCSLAPNIISQL
B18mouse    TCSLAPNILSQL
            ********.***
```

CLUSTAL V alignment of human and rat cDNA sequences (identity: 612/780, 78.5%):

```
B18hum      GCCAGGAGGCGGAGTGGAAGTGGCCGTGGGGCGGGTATGGGACTAGCTGG
B18rat      --------------TG--------CGTC--------------CTAGGCAG
                          **        ***               ****   *

B18hum      CGTGTGCGCCCTGAGACGCTCAGCGGGCTATATACTCGTCGGTGGGGCCG
B18rat      CATA---GCC---GGA---TCGGTGCTCCGTGCATCGGCTACTTGGAC--
            * *    ***    **   ** * *  *  *  *   *    * ** *

B18hum      GCGGTCAGTCTGCGGCAGCGGCAGCAAGACGGTGCAGTGAAGGAGAGTGG
B18rat      -------GTGCGTGGCAGGCAGAGCAGGCCGG---AAAGGAGCAGGTTGG
                   **  * *****    **** * ***   *  * ** **  ***
```

FIG.5b

```
B18hum          GCGTCTGGCGGGGTCCGCAGTTTCAGCAGAGCCGCTGCAGCCATGGCCCC
B18rat          GAGTGTGGTGGGGCCCGCAGCTTCAGCAGTGCCGCGGTGACTATGGCCCC
                * ** *** **** ****** ******** ***** *   * ********

B18hum          AATCAAGGTGGGAGATGCCATCCCAGCAGTGGAGGTGTTTGAAGGGGAGC
B18rat          GATCAAGGTGGGAGACACCATTCCCTCAGTGGAGGTATTTGAAGGGGAAC
                 **************  **** **  ********** *********** *

B18hum          CAGGGAACAAGGTGAACCTGGCAGAGCTGTTCAAGGGCAAGAAGGGTGTG
B18rat          CTGGAAAGAAGGTGAACTTGGCAGAGCTGTTCAAGGACAAGAAAGGTGTT
                * ** ** ********* ****************** ****** ***** 

B18hum          CTGTTTGGAGTTCCTGGGGCCTTCACCCCTGGATGTTCCAAGACACACCT
B18rat          TTGTTTGGAGTCCCTGGGGCATTTACACCTGGCTGTTCCAAGACCCATCT
                 ********** ******** ** ** ***** *********** ** **

B18hum          GCCAGGGTTTGTGGAGCAGGCTGAGGCTCTGAAGGCCAAGGGAGTCCAGG
B18rat          GCCTGGGTTTGTGGAGCAAGCCGGAGCTCTGAAGGCCAAGGGAGCACAAG
                *** ************** ** *  *******************  ** *

B18hum          TGGTGGCCTGTCTGAGTGTTAATGATGCCTTTGTGACTGGCGAGTGGGGC
B18rat          TGGTGGCCTGTCTGAGTGTTAATGATGTCTTCGTGACTGCAGAGTGGGGT
                *************************** *** *******  ******** 

B18hum          CGAGCCCACAAGGCGGAAGGCAAGGTTCGGCTCCTGGCTGATCCCACTGG
B18rat          CGAGCCCACCAGGCAGAAGGCAAGGTTCAGCTCCTGGCTGACCCCACTGG
                ********* **** ************* ************ ********

B18hum          GGCCTTTGGGAAGGAGACAGACTTATTACTAGATGATTCGCTGGTGTCCA
B18rat          AGCTTTTGGAAAGGAGACAGATTTACTACTAGATGATTCTTTGGTGTCTC
                 ** ***** *********** *** *************  *******  

B18hum          TCTTTGGGAATCGACGTCTCAAGAGGTTCTCCATGGTGGTACAGGATGGC
B18rat          TCTTTGGGAATCGTCGGCTAAAAAGGTTCTCCATGGTGATAGACAAGGGC
                ************* ** ** ** *************** ** *  * ***

B18hum          ATAGTGAAGGCCCTGAATGTGGAACCAGATGGCACAGGCCTCACCTGCAG
B18rat          GTAGTAAAGGCACTGAACGTGGAGCCGGATGGCACAGGCCTCACCTGCAG
                 **** ***** ***** ***** ** ***********************

B18hum          CCTGGCACCCAATATCATCTCACAGCTCTGAGGCCCTGGGCCAGATTACT
B18rat          CCTGGCCCCCAACATCCTCTCACAACTCTGAGGCCCTGA-CCAGA--ATG
                ****** ***** *** ******* *************  *****  *  

B18hum          TCCTCCACCCCTCCCTATCTCACCTGCCCAGCCCTGTGCTGG-GGCCCTG
B18rat          TCCTCTGACTCTCCC-ATCTCCTCCACCCAGCTCTGGGCCAAAGGCCCAG
                *****   * ***** *****  *  ****** *** **    ***** *

B18hum          CA------------------ATTGGAATG-----TTGGCCAGATTTCTGC
B18rat          TACCTCCTTACCTGAGGGCCACTGGAATGGAACCTTGACAATATTTCTGC
                 *                  * *******     *** * * ********

B18hum          AATAAACACTTGTGGTTTGCGGAAAAAAA-------------
B18rat          AATAAACAGTT-TAATTTGTGAAAAAAAAAAAAAAAAAAAAA
                ******** ** *  **** * *******
```

CLUSTAL V alignment of human and mouse cDNA sequences (Identity: 552/675, 81.8%):

FIG.5c

```
B18hum          GCCAGGAGGCGGAGTGGAAGTGGCCGTGGGGCGGGTATGGACTAGCTGG
B18mouse        -------------------------------------------------

B18hum          CGTGTGCGCCCTGAGACGCTCAGCGGGCTATATACTCGTCGGTGGGGCCG
B18mouse        ----------------TGCTCCGTG-----------CATCGACGTGCTTG
                                 **** * *           * ***  * *   *

B18hum          GCGGTCAGTCTGCGGCAGCGGCAGCAAGACGGTGCAGTGAAGGAGAGTGG
B18mouse        GCAGGCAG--------------AGCAGGCCGG---AAAGAAGCAGGTTGG
                ** * ***              **** * ***   *  **** **  ***

B18hum          GCGTCTGGCGGGGTCCGCAGTTTCAGCAGAGCCGCTGCAGCCATGGCCCC
B18mouse        GAGTGTGGCGGAGCCCGCAGCTTCAGCAGCTCCGCGGTGACCATGGCCCC
                * ** ****** * ****** ********  **** *  ***********

B18hum          AATCAAGGTGGGAGATGCCATCCCAGCAGTGGAGGTGTTTGAAGGGGAGC
B18mouse        GATCAAGGTGGGAGATGCCATTCCCTCAGTGGAGGTATTTGAAGGGGAAC
                 ******************** **  ********** *********** *

B18hum          CAGGGAACAAGGTGAACCTGGCAGAGCTGTTCAAGGGCAAGAAGGGTGTG
B18mouse        CGGGAAAGAAGGTGAACTTGGCAGAGCTGTTCAAGGGCAAGAAAGGTGTT
                * ** ** ********* ************************* *****

B18hum          CTGTTTGGAGTTCCTGGGGCCTTCACCCCTGGATGTTCCAAGACACACCT
B18mouse        TTGTTTGGAGTCCCTGGGGCATTTACACCTGGCTGTTCTAAGACCCACCT
                 ********** ******** ** ** ***** ***** ***** *****

B18hum          GCCAGGGTTTGTGGAGCAGGCTGAGGCTCTGAAGGCCAAGGGAGTCCAGG
B18mouse        GCCTGGGTTTGTGGAGCAAGCTGGAGCTCTGAAGGCTAAGGGAGCGCAGG
                *** ************** ****  *********** *******  ****

B18hum          TGGTGGCCTGTCTGAGTGTTAATGATGCCTTTGTGACTGGCGAGTGGGGC
B18mouse        TGGTGGCCTGTCTGAGCGTTAATGACGTCTTTGTGATTGAAGAGTGGGGT
                **************** ******** * ******** **  ********

B18hum          CGAGCCCACAAGGCGGAAGGCAAGGTTCGGCTCCTGGCTGATCCCACTGG
B18mouse        CGAGCCCACCAGGCAGAAGGCAAGGTTCGGCTCCTGGCTGACCCCACTGG
                ********* **** ************************** ********

B18hum          GGCCTTTGGGAAGGAGACAGACTTATTACTAGATGATTCGCTGGTGTCCA
B18mouse        AGCCTTTGGGAAGGCGACAGACTTATTATTGGATGATTCTTTGGTGTCTC
                 ************* ************* * ********  *******

B18hum          TCTTTGGGAATCGACGTCTCAAGAGGTTCTCCATGGTGGTACAGGATGGC
B18mouse        TCTTTGGGAATCGTCGGCTGAAAAGGTTCTCCATGGTGATAGACAACGGC
                ************* ** ** ** *************** ** *  * ***

B18hum          ATAGTGAAGGCCCTGAATGTGGAACCAGATGGCACAGGCCTCACCTGCAG
B18mouse        ATAGTGAAGGCACTGAACGTGGAGCCAGATGGCACAGGCCTCACCTGCAG
                *********** ***** ***** **************************

B18hum          CCTGGCACCCAATATCATCTCACAGCTCTGAGGCCCTGGGCCAGATTACT
B18mouse        CCTGGCCCCCAACATCCTCTCCCAACTCTGAGGCCCTGG-CCAGATG---
                ****** ***** *** **** ** ************** ******

B18hum          TCCTCCACCCCTCCCTATCTCACCTGCCCAGCCCTGTGCTGGGGCCCTGC
B18mouse        TCCTCTGACTCTCCC-ATCTCTCCCACCCGGCTCT------AGGCC----
                *****   * ***** ***** **  *** ** **      ****

B18hum          AATTGGAATGTTGGCCAGATTTCTGCAATAAACACTTGTGGTTTGCGGAA
B18mouse        ----AAAAGGCTCGGTACCTCCTTACTGGGAGC-CACGT-----------
                      ** * * * **  * * * *      * *  **
```

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: UNIVERSITE CATHOLIQUE DE LOUVAIN
Halles Universitaires
(B) STREET: Place de l' Universite, 1
(C) CITY: LOUVAIN-LA-NEUVE
(E) COUNTRY: BELGIUM
(F) POSTAL CODE (ZIP): B-1348

(A) NAME: UNIVERSITE DE MONS-HAINAUT
(B) STREET: Place du Parc 20
(C) CITY: MONS
(E) COUNTRY: BELGIUM
(F) POSTAL CODE (ZIP): B-7000

(ii) TITLE OF INVENTION: PEROXISOME-ASSOCIATED PEPTIDE, NUCLEOTIDE SEQUENCE ENCODING SAID PEPTIDE AND THEIR USES IN THE DIAGNOSTIC AND/OR THE TREATMENT OF LUNG INJURIES AND DISEASES, AND OF OXIDATIVE STRESS-RELATED DISORDERS

(iii) NUMBER OF SEQUENCES: 19

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 805 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:193..681

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GCCAGGAGGC GGAGTGGAAG TGGCCGTGGG GCGGGTATGG GACTAGCTGG CGTGTGCGCC 60

CTGAGACGCT CAGCGGGCTA TATACTCGTC GGTGGGGCCG GCGGTCAGTC TGCGGCAGCG 120

```
GCAGCAAGAC GGTGCAGTGA AGGAGAGTGG GCGTCTGGCG GGGTCCGCAG TTTCAGCAGA      180

GCCGCTGCAG CC ATG GCC CCA ATC AAG GTG GGA GAT GCC ATC CCA GCA          228
              Met Ala Pro Ile Lys Val Gly Asp Ala Ile Pro Ala
               1               5                  10

GTG GAG GTG TTT GAA GGG GAG CCA GGG AAC AAG GTG AAC CTG GCA GAG        276
Val Glu Val Phe Glu Gly Glu Pro Gly Asn Lys Val Asn Leu Ala Glu
         15                  20                  25

CTG TTC AAG GGC AAG AAG GGT GTG CTG TTT GGA GTT CCT GGG GCC TTC        324
Leu Phe Lys Gly Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe
     30                  35                  40

ACC CCT GGA TGT TCC AAG ACA CAC CTG CCA GGG TTT GTG GAG CAG GCT        372
Thr Pro Gly Cys Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala
 45                  50                  55                  60

GAG GCT CTG AAG GCC AAG GGA GTC CAG GTG GTG GCC TGT CTG AGT GTT        420
Glu Ala Leu Lys Ala Lys Gly Val Gln Val Val Ala Cys Leu Ser Val
                 65                  70                  75

AAT GAT GCC TTT GTG ACT GGC GAG TGG GGC CGA GCC CAC AAG GCG GAA        468
Asn Asp Ala Phe Val Thr Gly Glu Trp Gly Arg Ala His Lys Ala Glu
             80                  85                  90

GGC AAG GTT CGG CTC CTG GCT GAT CCC ACT GGG GCC TTT GGG AAG GAG        516
Gly Lys Val Arg Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Glu
         95                 100                 105

ACA GAC TTA TTA CTA GAT GAT TCG CTG GTG TCC ATC TTT GGG AAT CGA        564
Thr Asp Leu Leu Leu Asp Asp Ser Leu Val Ser Ile Phe Gly Asn Arg
    110                 115                 120

CGT CTC AAG AGG TTC TCC ATG GTG GTA CAG GAT GGC ATA GTG AAG GCC        612
Arg Leu Lys Arg Phe Ser Met Val Val Gln Asp Gly Ile Val Lys Ala
125                 130                 135                 140

CTG AAT GTG GAA CCA GAT GGC ACA GGC CTC ACC TGC AGC CTG GCA CCC        660
Leu Asn Val Glu Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro
                145                 150                 155

AAT ATC ATC TCA CAG CTC TGA GGCCCTGGGC CAGATTACTT CCTCCACCCC          711
Asn Ile Ile Ser Gln Leu  *
            160

TCCCTATCTC ACCTGCCCAG CCCTGTGCTG GGGCCCTGCA ATTGGAATGT TGGCCAGATT      771

TCTGCAATAA ACACTTGTGG TTTGCGGAAA AAAA                                  805
```

```
(2) INFORMATION FOR SEQ ID NO: 2:

     (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 163 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
```

```
Met Ala Pro Ile Lys Val Gly Asp Ala Ile Pro Ala Val Glu Val Phe
1               5                   10                  15

Glu Gly Glu Pro Gly Asn Lys Val Asn Leu Ala Glu Leu Phe Lys Gly
            20                  25                  30

Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe Thr Pro Gly Cys
        35                  40                  45

Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala Glu Ala Leu Lys
    50                  55                  60

Ala Lys Gly Val Gln Val Val Ala Cys Leu Ser Val Asn Asp Ala Phe
65                  70                  75                  80

Val Thr Gly Glu Trp Gly Arg Ala His Lys Ala Glu Gly Lys Val Arg
                85                  90                  95

Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Glu Thr Asp Leu Leu
            100                 105                 110

Leu Asp Asp Ser Leu Val Ser Ile Phe Gly Asn Arg Arg Leu Lys Arg
        115                 120                 125

Phe Ser Met Val Val Gln Asp Gly Ile Val Lys Ala Leu Asn Val Glu
    130                 135                 140

Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro Asn Ile Ile Ser
145                 150                 155                 160

Gln Leu  *
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 780 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Rattus Rattus

(ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION:136..624

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
TGCGTCCTAG GCAGCATAGC CGGATCGGTG CTCCGTGCAT CGGCTACTTG GACGTGCGTG      60

GCAGGCAGAG CAGGCCGGAA AGGAGCAGGT TGGGAGTGTG GTGGGGCCCG CAGCTTCAGC     120
```

AGTGCCGCGG TGACTATGGC CCCGATCAAG GTGGGAGACA CCATTCCCTC AGTGGAGGTA 180

TTTGRAGGGG AACCTGGAAA GAAGGTGAAC TTGGCAGAGC TGTTCAAGGA CAAGAAAGGT 240

GTTTTGTTTG GAGTCCCTGG GGCATTTACA CCTGGCTGTT CCAAGACCCA TCTGCCTGGG 300

TTTGTGGAGC AAGCCGGAGC TCYGAAGGCC AAGGGAGCAC AAGTGGTGGC CTGTCTGAGT 360

GTTAATGATG YCTTCGTGAC TGCAGAGTGG GGTCGAGCCC ACCAGGCAGA AGGCAAGGTT 420

CAGCTCCTGG CTGACCCCAC TGGAGCTTTT GGAAAGGAGA CAGATTTACT ACTAGATGAT 480

TCTTTGGTGT CTCTCTTTGG GAATCGTCGG CTAAAAAGGT TCTCCATGGT GATAGACAAG 540

GGCGTAGTAA AGGCACTGAA CGTGGAGCCG GATGGCACAG GCCTCACCTG CAGCCTGGCC 600

CCCAACATCC TCTCACAACT CTGAGGCCCT GACCAGAATG TCCTCTGACT CTCCCATCTC 660

CTCCACCCAG CTCTGGGCCA AAGGCCCAGT ACCTCCTTAC CTGAGGGCCA CTGGAATGGA 720

ACCTTGACAA TATTTCTGCA ATAAACAGTT TAATTTGTGA AAAAAAAAAA AAAAAAAAAA 780

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 162 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus Rattus

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION:17
(D) OTHER INFORMATION:/product= "Glu or Gly"

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION:63
(D) OTHER INFORMATION:/product= "Leu or Pro"

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION:79
(D) OTHER INFORMATION:/product= "Ala or Val"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Ala Pro Ile Lys Val Gly Asp Thr Ile Pro Ser Val Glu Val Phe
1               5                   10                  15
```

```
Xaa Gly Glu Pro Gly Lys Lys Val Asn Leu Ala Glu Leu Phe Lys Asp
            20                  25                  30

Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe Thr Pro Gly Cys
        35                  40                  45

Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala Gly Ala Xaa Lys
    50                  55                  60

Ala Lys Gly Ala Gln Val Val Ala Cys Leu Ser Val Asn Asp Xaa Phe
65                  70                  75                  80

Val Thr Ala Glu Trp Gly Arg Ala His Gln Ala Glu Gly Lys Val Gln
                85                  90                  95

Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Glu Thr Asp Leu Leu
            100                 105                 110

Leu Asp Asp Ser Leu Val Ser Leu Phe Gly Asn Arg Arg Leu Lys Arg
        115                 120                 125

Phe Ser Met Val Ile Asp Lys Gly Val Val Lys Ala Leu Asn Val Glu
    130                 135                 140

Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro Asn Ile Leu Ser
145                 150                 155                 160

Gln Leu
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 675 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mouse

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:99..588

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
TGCTCCGTGC ATCGACGTGC TTGGCAGGCA GAGCAGGCCG GAAAGAAGCA GGTTGGGAGT     60

GTGGCGGAGC CCGCAGCTTC AGCAGCTCCG CGGTGACCAT GGCCCCGATC AAGGTGGGAG    120

ATGCCATTCC CTCAGTGGAG GTATTTGAAG GGGAACCGGG AAAGAAGGTG AACTTGGCAG    180

AGCTGTTCAA GGGCAAGAAA GGTGTTTTGT TTGGAGTCCC TGGGGCATTT ACACCTGGCT    240
```

```
GTTCTAAGAC CCACCTGCCT GGGTTTGTGG AGCAAGCTGG AGCTCTGAAG GCTAAGGGAG    300

CGCAGGTGGT GGCCTGTCTG AGCGTTAATG ACGTCTTTGT GATTGAAGAG TGGGGTCGAG    360

CCCACCAGGC AGAAGGCAAG GTTCGGCTCC TGGCTGACCC CACTGGAGCC TTTGGGAAGG    420

CGACAGACTT ATTATTGGAT GATTCTTTGG TGTCTCTCTT TGGGAATCGT CGGCTGAAAA    480

GGTTCTCCAT GGTGATAGAG AACGGCATAG TGAAGGCACT GAACGTGGAG CCAGATGGCA    540

CAGGCCTCAC CTGCAGCCTG GCCCCCAACA TCCTCTCCCA ACTCTGAGGC CCTGGCCAGA    600

TGTCCTCTGA CTCTCCCATC TCTCCCACCC GGCTCTAGGC CAAAAGGCTC GGTACCTCCT    660

TACTGGGAGC CACGT                                                     675
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 162 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mouse

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Ala Pro Ile Lys Val Gly Asp Ala Ile Pro Ser Val Glu Val Phe
1               5                   10                  15

Glu Gly Glu Pro Gly Lys Lys Val Asn Leu Ala Glu Leu Phe Lys Gly
            20                  25                  30

Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe Thr Pro Gly Cys
        35                  40                  45

Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala Gly Ala Leu Lys
    50                  55                  60

Ala Lys Gly Ala Gln Val Val Ala Cys Leu Ser Val Asn Asp Val Phe
65                  70                  75                  80

Val Ile Glu Glu Trp Gly Arg Ala His Gln Ala Glu Gly Lys Val Arg
                85                  90                  95

Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Ala Thr Asp Leu Leu
            100                 105                 110

Leu Asp Asp Ser Leu Val Ser Leu Phe Gly Asn Arg Arg Leu Lys Arg
        115                 120                 125
```

```
Phe Ser Met Val Ile Asp Asn Gly Ile Val Lys Ala Leu Asn Val Glu
    130                 135                 140

Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro Asn Ile Leu Ser
145                 150                 155                 160

Gln Leu
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 469 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:161..382

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
GGGTATGGGA CTAGCTGGCG TGTGCGCCCT GAGACGCTCA GCGGGCTATA TACTCGTCGG      60

TGGGGCCGGC GGTCAGTCTG CGGCAGCGGC AGCAAGACGG TGCAGTGAAG GAGAGTGGGC     120

GTCTGGCGGG GTCCGCAGTT TCAGCAGAGC CGCTGCAGCC ATGGCCCCAA TCAAGGTTCG     180

GCTCCTGGCT GATCCCACTG GGGCCTTTGG GAAGGAGACA GACTTATTAC TAGATGATTC     240

GCTGGTGTCC ATCTTTGGGA ATCGACGTCT CAAGAGGTTC TCCATGGTGG TACAGGATGG     300

CATAGTGAAG GCCCTGAATG TGGAACCAGA TGGCACAGGC CTCACCTGCA GCCTGGCACC     360

CAATATCATC TCACAGCTCT GAGGCCCTGG GCCAGATTAC TTCCTCCACC CCTCCCTATC     420

TCACCTGCCC AGCCGTGTGC TGGGGCCCTG CAATTGGAAT GTTGGCCAG                 469
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:161..514

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
GGGTATGGGA CTAGCTGGCG TGTGCGCCCT GAGACGCTCA GCGGGCTATA TACTCGTCGG    60

TGGGGCCGGC GGTCAGTCTG CGGCAGCGGC AGCAAGACGG TGCAGTGAAG GAGAGTGGGC   120

GTCTGGCGGG GTCCGCAGTT TCAGCAGAGC CGCTGCAGCC ATGGCCCCAA TCAAGACACA   180

CCTGCCAGGG TTTGTGGAGC AGGCTGAGGC TCTGAAGGCC AAGGGAGTCC AGGTGGTGGC   240

CTGTCTGAGT GTTAATGATG CCTTTGTGAC TGGCGAGTGG GGCCGAGCCC ACAAGGCGGA   300

AGGCAAGGTT CGGCTCCTGG CTGATCCCAC TGGGGCCTTT GGGAAGGAGA CAGACTTATT   360

ACTAGATGAT TCGCTGGTGT CCATCTTTGG GAATCGACGT CTCAAGAGGT TCTCCATGGT   420

GGTACAGGAT GGCATAGTGA AGGCCCTGAA TGTGGAACCA GATGGCACAG GCCTCACCTG   480

CAGCCTGGCA CCCAATATCA TCTCACAGCT CTGAGGCCCT GGGCCAGATT ACTTCCTCCA   540

CCCCTCCCTA TCTCACCTGC CCAGCCCTGT GCTGGGGCCC TGCAATTGGA ATGTTGGCCA   600

G                                                                   601
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 604 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:161..517

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
GGGTATGGGA CTAGCTGGCG TGTGCGCCCT GAGACGCTCA GCGGGCTATA TACTCGTCGG    60
```

```
TGGGGCCGGC GGTCAGTCTG CGGCAGCGGC AGCAAGACGG TGCAGTGAAG GAGAGTGGGC     120
GTCTGGCGGG GTCCGCAGTT TCAGCAGAGC CGCTGCAGCC ATGGCCCCAA TCAAGGTGGG     180
AGATGCCATC CCAGCAGTGG AGGTGTTTGA AGGGGAGCCA GGGAACAAGG TGAACCTGGC     240
AGAGCTGTTC AAGGGCAAGA AGGTGTGCT  GTTTGGAGTT CCTGGGGCCT TCACCCCTGG     300
ATGTTCCAAG GTTCGGCTCC TGGCTGATCC CACTGGGGCC TTTGGGAAGG AGACAGACTT     360
ATTACTAGAT GATTCGCTGG TGTCCATCTT TGGGAATCGA CGTCTCAAGA GGTTCTCCAT     420
GGTGGTACAG GATGGCATAG TGAAGGCCCT GAATGTGGAA CCAGATGGCA CAGGCCTCAC     480
CTGCAGCCTG GCACCCAATA TCATCTCACA GCTCTGAGGC CCTGGGCCAG ATTACTTCCT     540
CCACCCCTCC CTATCTCACC TGCCCAGCCC TGTGCTGGGG CCCTGCAATT GGAATGTTGG     600
CCAG                                                                  604
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2710 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:2516..2710

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:2074..2135

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:1932..1970

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:1728..1859

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:802..936

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
TCTGTCCCTT AGCGCCCCG CGGGGGCTTA CCCCATCCCA CTCCATGACC TCCCCTCCCC     60
CCATGGCGAA TTCCCACCTT TCTGTCTTTC ACTCACTTCC TGGAACCGTC CCCAGGGCCT    120
TGGACCTTCC CCCTTCTCCT CCCAAACCTT GTGAGACCCC ATTCCCTTTC TACTTCATCC    180
TGCTCTCAAC TTTTGGGCTC CTCAGAGGCC CTCACCCCTG ACTCTCTCTC CCTACCCACT    240
CTGGTCCCAT GAAGCCCTCA AGTACTCTGG GGATGGATCC TTCCCCCTTC AAAAGATTCC    300
TTCTTTTGTT CTACACCTCC TGGGTGTAGG GGCCTGGACA CCCTCCCCCA ACGTTCCACC    360
TGCCGCTGCC CTTCCTCTTC CTCCTCCTGA GGGTGGGACC CTCAGACCTG GCCAAGATCC    420
TCTCCCTCCA TGTTGTCAGG GACTCCTCCT CACCCCCAAA TACAGCCCTC TAGCCCCTGT    480
CCATTTTATT CCACTCCTTT CCTGTAACCT AGACAGCATG TTATGCAACC CTTTGCGACA    540
CATGGGGAAA CCTTCCCTCC CTTCCTCTGT TGTCACCAAT GGCCCCTTAA GAGGAGCAGG    600
GCCACCTTGA AACTTGGAGG ATATGGGGTA ACCCAGTGGG AGCGGGCAGG GAGGGCCCTT    660
GGAAACTGAC AGGGCTGGAG TATCCTGCTG GGTTTCAGCC CCGGTTCCTG CAGGCACAGC    720
TGCCAGGCTC TCTGTTCACC TTCCTGCCTC TGGTTTGCCC CGGCTCCCTC ACCCCCCTTA    780
CCCTGGAGTC CTTCCTTCTA GGTGGGAGAT GCCATCCCAG CAGTGGAGGT GTTTGAAGGG    840
GAGCCAGGGA ACAAGGTGAA CCTGGCAGAG CTGTTCAAGG GCAAGAAGGG TGTGCTGTTT    900
GGAGTTCCTG GGGCCTTCAC CCCTGGATGT TCCAAGGTGA GGCCCTTCCC CTTCTGAAGA    960
TCAGGACCTG GGGATCTTTT GTGTTGCTCT TAAGTCCTCC ACATAGTCCT GATAGGACTC   1020
CTAAAAAGCA TTTCAGTGCC ATCACAAAAC AAGTAGAGCT GGGTAGAGCT GGGCGCGGTG   1080
GCTCACGCCT GTAATCCCAG CACTTTGGGA GGCCAAGGCG GGTGGATCAC GAGGTCAGGA   1140
GTCCAAAACC AGCCTGGCCA AGATGGTGAA ACCCTGTCTC TACTAAAAAT GCAAAAAAAT   1200
CAGCCGGATA TGGTGGCGGG CGCCTGTAAT CCCAGGTATT GGGGAGGCTG AGGCAGAGAA   1260
TTGCTTGAAC CCAGGAGGCG TAGGTTGCAG TGAGTGGAGA TCGTGCCTCT GCAGTCCAGC   1320
CTGGGTGAAA GAGCGAGACT CCGTCTCAAA ATGAAAAAAA AAAAAGAAAA CAAGTAGAGA   1380
CTGCAAAAAG GGAACAGTAC CGGGAATGTT GGAGAAAAAC ATACTACAAT TAAATCCAAC   1440
ACCCCTGTTG GTCCTGCTAA ATGACAGGCA CTGTGGAAGG TGCTTGGGAC TCAGATAAAT   1500
AAGACAAAGA TCTGCCCATG GAAAGTTCAC GTCTGGACCA TAAGGCATTA GGTTTCATTC   1560
TGAGCTTCCT AGTGGCCAAG GCAAAAAGGA AATAGAATGG TTTAGACAGC TCTCATTGTC   1620
TGATCAAAGG TGTTGAGGCA GAGCACTGAG GAGGGCCTGG AGATAAAGGG TGGGCTGGGG   1680
GTCAGATGCA GTTATCCCTT TGCCGACCCT TTGTTCCCCT TCCTCAGACA CACCTGCCAG   1740
GGTTTGTGGA GCAGGCTGAG GCTCTGAAGG CCAAGGGAGT CCAGGTGGTG GCCTGTCTGA   1800
GTGTTAATGA TGCCTTTGTG ACTGGCGAGT GGGGCCGAGC CCACAAGGCG GAAGGCAAGG   1860
```

TGAGGTGAGG GGCCTGCAGG GAGTCAGGAC CAGGTAGGAT ATTCTTCTTG TGACCTCTAC 1920

TTTCTCTGCA GGTTCGGCTC CTGGCTGATC CCACTGGGGC CTTTGGGAAG GTGAGTGTTC 1980

CCCTGACCGC CACAGGGACA TGGCGGTGCG GGGAGCAGTG GGGGCCCTTG GCCTCTTCAA 2040

GGATTTCTGA CACTTTTCTC TGTCTCTTCT TAGGAGACAG ACTTATTACT AGATGATTCG 2100

CTGGTGTCCA TCTTTGGGAA TCGACGTCTC AAGAGGTAAA AGTGGAGAGT CCTCTGTGGA 2160

GAAAGTCCTC TGTGGGAGAG AGTCCTCTGT GGGAGAGAGT CCTCTGTGGA GAGGGTCCTC 2220

TGTGGGAAGA GTCGTCTGTG GGGGAGATGT GTGGGAGAGA GTCCTGTGTG GGGAGAGTCT 2280

TCTGTAGGGG AGAGTCCTCT GGGGAGAGAG TCCTGTGTGG GGGAGAGTCC TCTGTGGGGA 2340

GAGTCCTCTG TGTGGAGAGA GTCCTGTGTG GTGGTGAGTC CTCTGTGGGG GAGAGTCCTC 2400

TGTGGGGGGA GTCCTCTCTG GAGTTCTCTT GGGCCCCTGG CTGTTCACTG CCTGTCTCCA 2460

TGCCCAGCCT CCAAGCCCAG GCTGATGCAG CTGGCTGGGC CCCTCTTTCC GGCAGGTTCT 2520

CCATGGTGGT ACAGGATGGC ATAGTGAAGG CCCTGAATGT GGAACCAGAT GGCACAGGCC 2580

TCACCTGCAG CCTGGCACCC AATATCATCT CACAGCTCTG AGGCCCTGGG CCAGATTACT 2640

TCCTCCACCC CTCCCTATCT CACCTGCCCA GCCCTGTGCT GGGGCCCTGC AATTGGAATG 2700

TTGGCCAGAT 2710

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GCCATCCCAG CAGTGGAGGT GTTTG 25

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TTGAACAGCT CTGCCAGGTT CACC 24

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TGGAGGTGTT TGAAGGGGAG CCAG 24

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CAGGTTCACC TTGTTCCCTG GCTC 24

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GGGTATGGGA CTAGCTGGCG 20

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CTGGCCAACA TTCCAATTGC AG 22

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATGTTATGCA ACCCTTTGCG ACAC 24

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GTGTTTGAAG GGGAGCCAGG GAAC 24

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGAGACAGGG TTTCACCATC TTGG 24

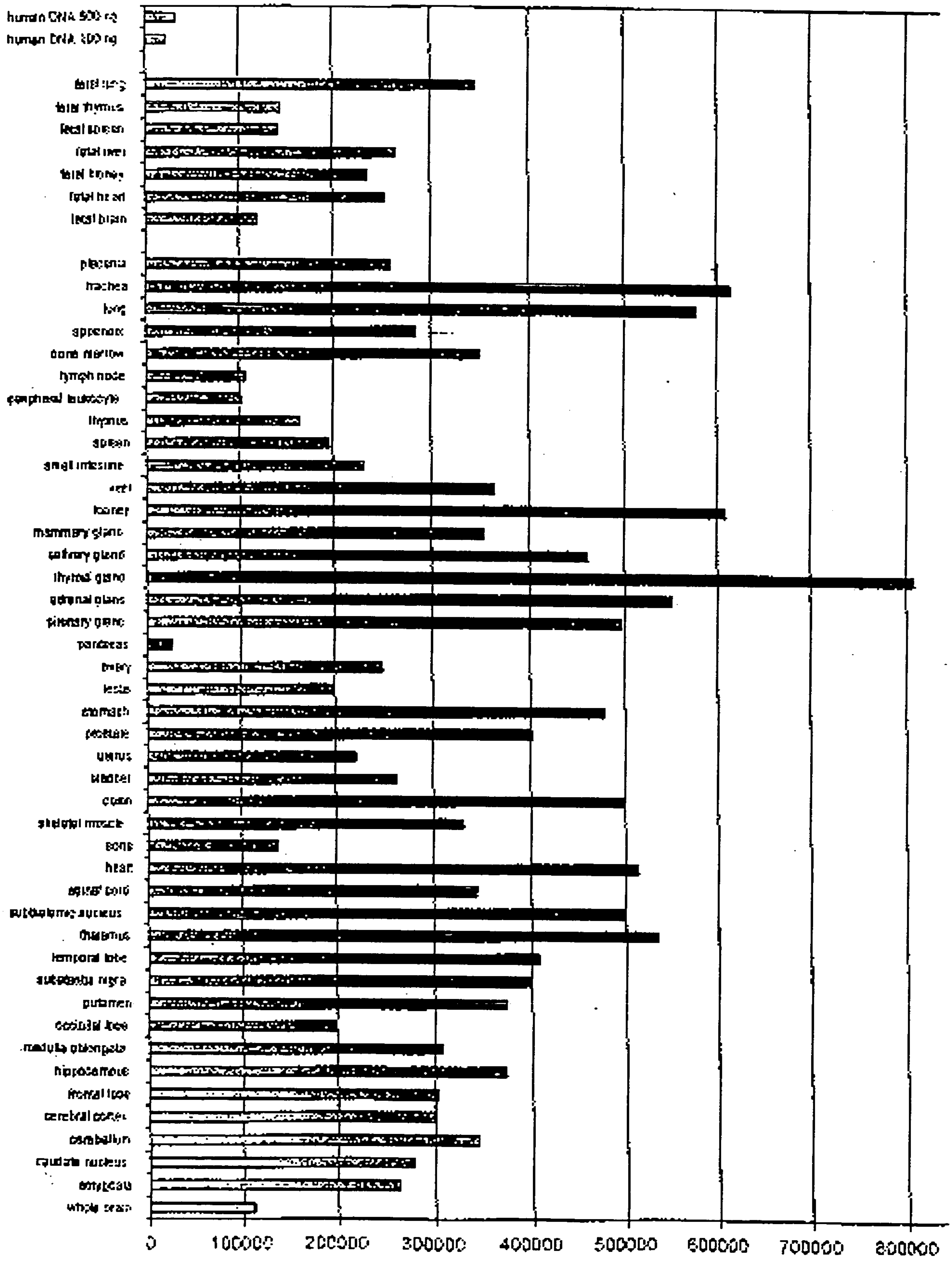

human DNA 500 ng
human DNA 100 ng
fetal lung
fetal thymus
fetal spleen
fetal liver
fetal kidney
fetal heart
fetal brain
placenta
trachea
lung
appendix
bone marrow
lymph node
peripheral leukocyte
thymus
spleen
small intestine
liver
kidney
mammary gland
salivary gland
thyroid gland
adrenal gland
pituitary gland
pancreas
ovary
testis
stomach
prostate
uterus
bladder
colon
skeletal muscle
aorta
heart
spinal cord
subthalamic nucleus
thalamus
temporal lobe
substantia nigra
putamen
occipital lobe
medulla oblongata
hippocampus
frontal lobe
cerebral cortex
cerebellum
caudate nucleus
amygdala
whole brain
0
100000
200000
300000
400000
500000
600000
700000
800000
B18 mRNA (arbitrary units)

FIG.1

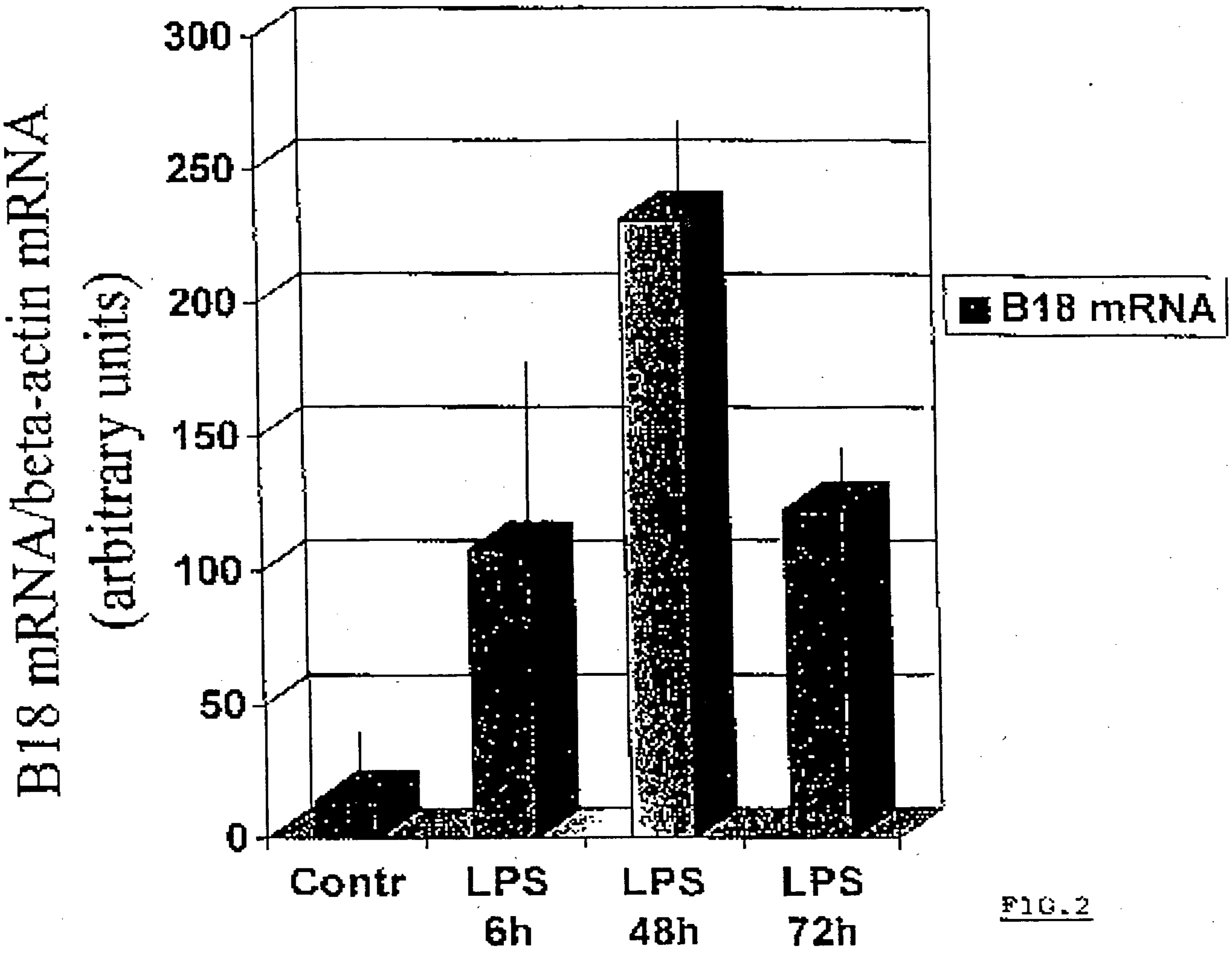

B18 mRNA/beta-actin mRNA
(arbitrary units)
300
250
200
150
100
50
0
■ B18 mRNA
Contr
LPS 6h
LPS 48h
LPS 72h

FIG.2

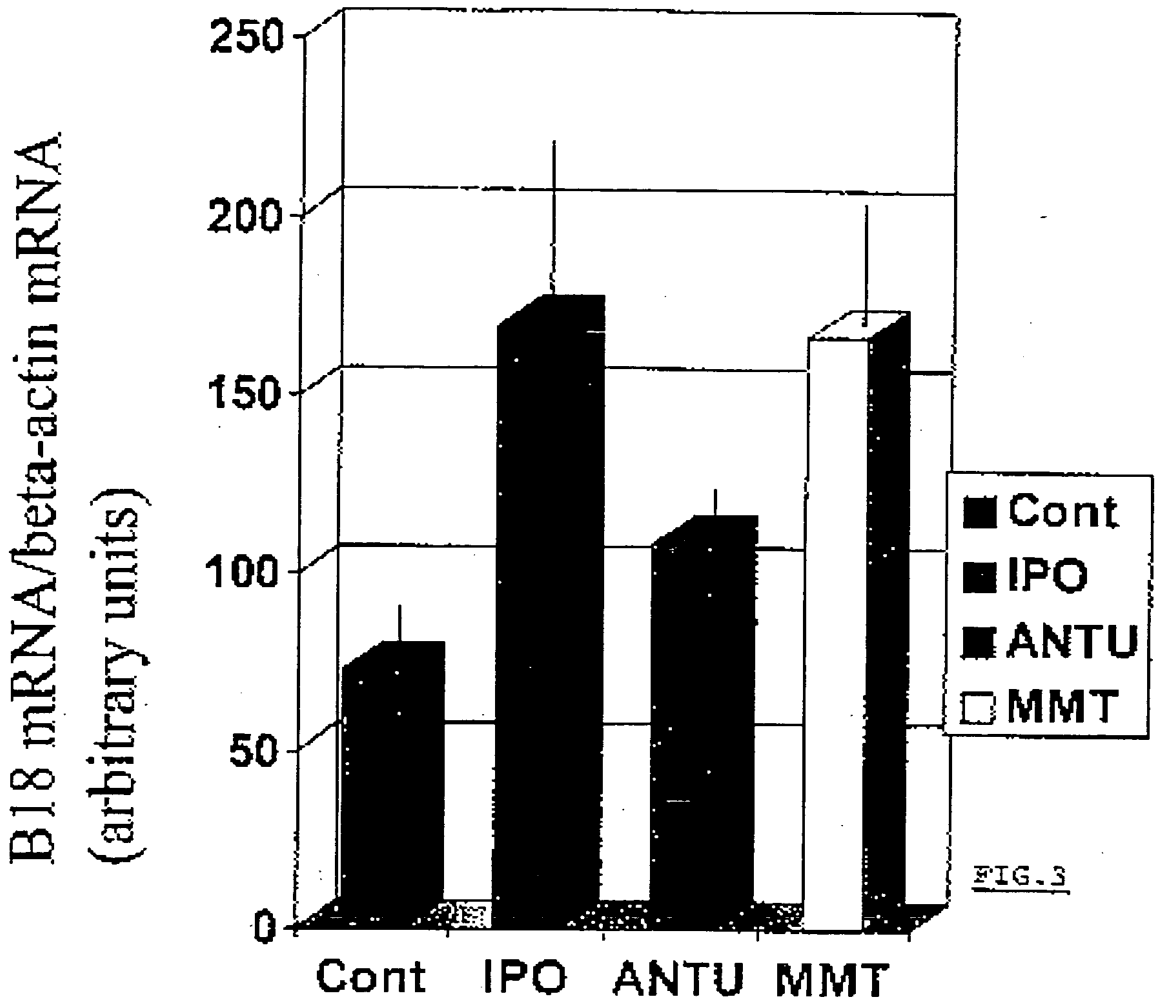

B18 mRNA/beta-actin mRNA
(arbitrary units)
250
200
150
100
50
0
Cont
IPO
ANTU
MMT
■ Cont
■ IPO
■ ANTU
□ MMT

FIG.3

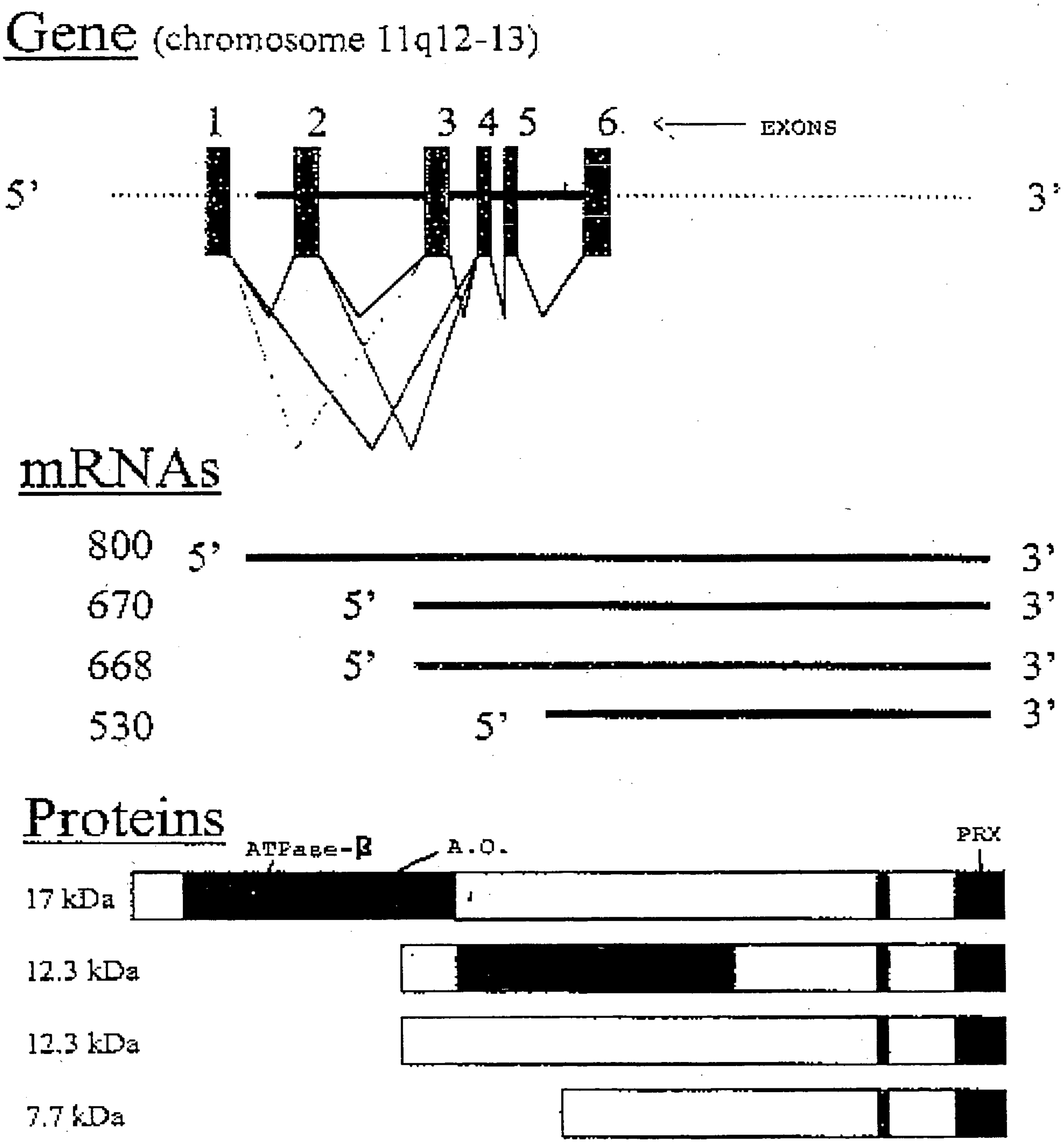

Gene (chromosome 11q12-13)
1
2
3
4
5
6.
EXONS
5'
3'
mRNAs
800
5'
3'
670
5'
3'
668
5'
3'
530
5'
3'
Proteins
ATPase-β
A.O.
PRX
17 kDa
12.3 kDa
12.3 kDa
7.7 kDa

FIG.4

CLUSTAL V alignment of human and rat B18 amino acid sequences (Identity: 90%, Homology: 97.5%):

```
B18hum       MAPIKVGDAIPAVEVFEGEPGNKVNLAELFKGKKGVLFGVPGAFTPGCSK = SEQIDNO1
B18rat       MAPIKVGDTIPSVEVFEGEPGKKVNLAELFKDKKGVLFGVPGAFTPGCSK
             ********.**.*********.*********.******************

B18hum       THLPGFVEQAEALKAKGVQVVACLSVNDAFVTGEWGRAHKAEGKVRLLAD
B18rat       THLPGFVEQAGALKAKGAQVVACLSVNDVFVTAEWGRAHQAEGKVQLLAD    FIG.5a
             ********** ****** ********** ***.******.*****.****

B18hum       PTGAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQDGIVKALNVEPDGTGL
B18rat       PTGAFGKETDLLLDDSLVSLFGNRRLKRFSMVIDKGVVKALNVEPDGTGL
             *******************.************.. *.*************

B18hum       TCSLAPNIISQL
B18rat       TCSLAPNILSQL
             ********.***
```

CLUSTAL V alignment of human and mouse B18 amino acid sequences (Identity: 91%, Homology: 96%):

```
B18hum       MAPIKVGDAIPAVEVFEGEPGNKVNLAELFKGKKGVLFGVPGAFTPGCSK
B18mouse     MAPIKVGDAIPSVEVFEGEPGKKVNLAELFKGKKGVLFGVPGAFTPGCSK
             ***********.*********.****************************

B18hum       THLPGFVEQAEALKAKGVQVVACLSVNDAFVTGEWGRAHKAEGKVRLLAD
B18mouse     THLPGFVEQAGALKAKGAQVVACLSVNDVFVIEEWGRAHQAEGKVRLLAD
             ********** ****** ********** **  ******.**********

B18hum       PTGAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQDGIVKALNVEPDGTGL
B18mouse     PTGAFGKATDLLLDDSLVSLFGNRRLKRFSMVIDNGIVKALNVEPDGTGL
             ******* ***********.************...***************

B18hum       TCSLAPNIISQL
B18mouse     TCSLAPNILSQL
             ********.***
```

CLUSTAL V alignment of human and rat cDNA sequences (identity: 612/780, 78.5%):

```
B18hum       GCCAGGAGGCGGAGTGGAAGTGGCCGTGGGGCGGGTATGGGACTAGCTGG
B18rat       --------------TG--------CGTC--------------CTAGGCAG
                           **        ***               ****   *

B18hum       CGTGTGCGCCCTGAGACGCTCAGCGGGCTATATACTCGTCGGTGGGGCCG
B18rat       CATA---GCC---GGA---TCGGTGCTCCGTGCATCGGCTACTTGGAC--
             * *    ***    **   ** * *  *  *  *   *    * ** *

B18hum       GCGGTCAGTCTGCGGCAGCGGCAGCAAGACGGTGCAGTGAAGGAGAGTGG
B18rat       -------GTGCGTGGCAGGCAGAGCAGGCCGG---AAAGGAGCAGGTTGG
                    **  * *****    **** * ***   *  * ** **  ***
```

FIG.5b

```
B18hum      GCGTCTGGCGGGGTCCGCAGTTTCAGCAGAGCCGCTGCAGCCATGGCCCC
B18rat      GAGTGTGGTGGGGCCCGCAGCTTCAGCAGTGCCGCGGTGACTATGGCCCC
            * ** *** **** ****** ******** ***** *   * ********

B18hum      AATCAAGGTGGGAGATGCCATCCCAGCAGTGGAGGTGTTTGAAGGGGAGC
B18rat      GATCAAGGTGGCAGACACCATTCCCTCAGTGGAGGTATTTGAAGGGGAAC
             **************  **** **  ********** *********** *

B18hum      CAGGGAACAAGGTGAACCTGGCAGAGCTGTTCAAGGCCAAGAAGGGTGTG
B18rat      CTGGAAAGAAGGTGAACTTGGCAGAGCTGTTCAAGGACAAGAAAGGTGTT
            * ** ** ********* ****************** ****** *****

B18hum      CTGTTTGGAGTTCCTGGGGCCTTCACCCCTGGATGTTCCAAGACACACCT
B18rat      TTGTTTGGAGTCCCTGGGGCATTTACACCTGGCTGTTCCAAGACCCATCT
             ********** ******** ** ** ***** *********** ** **

B18hum      GCCAGGGTTTGTGGAGCAGGCTGAGGCTCTGAAGGCCAAGGGAGTCCAGG
B18rat      GCCTGGGTTTGTGGAGCAAGCCGGAGCTCTGAAGGCCAAGGGAGCACAAG
            *** ************** ** *  *******************  ** *

B18hum      TGGTGGCCTGTCTGAGTGTTAATGATGCCTTTGTGACTGGCGAGTGGGGC
B18rat      TGGTGGCCTGTCTGAGTCTTAATGATGTCTTCGTGACTGCAGAGTGGGGT
            *************************** *** *******  ********

B18hum      CGAGCCCACAAGGCGGAAGGCAAGGTTCGGCTCCTGGCTGATCCCACTGG
B18rat      CGAGCCCACCAGGCAGAAGGCAAGGTTCAGCTCCTGGCTGACCCCACTGG
            ********* **** ************* ************* *******

B18hum      GGCCTTTGGGAAGGAGACAGACTTATTACTAGATGATTCGCTGGTGTCCA
B18rat      AGCTTTTGGAAAGGAGACAGATTTACTACTAGATGATTCTTTGGTGTCTC
             ** ***** *********** *** *************  *******

B18hum      TCTTTGGGAATCGACGTCTCAAGACGTTCTCCATGGTGGTACAGGATGGC
B18rat      TCTTTGGGAATCGTCGGCTAAAAAGGTTCTCCATGGTGATAGACAAGGGC
            ************* ** ** ** * ************* ** *  * ***

B18hum      ATAGTGAAGGCCCTGAATGTGGAACCAGATGGCACAGGCCTCACCTGCAG
B18rat      GTAGTAAAGGCACTGAACGTGGAGCCGGATGGCACAGGCCTCACCTGCAG
             **** ***** ***** ***** ** ***********************

B18hum      CCTGGCACCCAATATCATCTCACAGCTCTGAGGCCCTGGGCCAGATTACT
B18rat      CCTGGCCCCCAACATCCTCTCACAACTCTGAGGCCCTGA-CCAGA--ATG
            ****** ***** *** ******* *************  *****  *

B18hum      TCCTCCACCCCTCCCTATCTCACCTGCCCAGCCCTGTGCTGG-GGCCCTG
B18rat      TCCTCTGACTCTCCC-ATCTCCTCCACCCAGCTCTGCGCCAAAGGCCCAG
            *****   * ***** ***** *  ****** *** **    ***** *

B18hum      CA------------------ATTGGAATG-----TTGGCCAGATTTCTGC
B18rat      TACCTCCTTACCTGAGGGCCACTGGAATGGAACCTTGACAATATTTCTGC
             *                  * *******     *** * * ********

B18hum      AATAAACACTTGTGGTTTGCGGAAAAAAA-------------
B18rat      AATAAACAGTT-TAATTTGTGAAAAAAAAAAAAAAAAAAAAA
            ******** ** *  **** * *******
```

CLUSTAL V alignment of human and mouse cDNA sequences (Identity: 552/675, 81.8%):

FIG.5c

```
B18hum           GCCAGGAGGCGGAGTGGAAGTGGCCGTGGGGCGGGTATGGGACTAGCTGG
B18mouse         --------------------------------------------------


B18hum           CGTGTGCGCCCTGAGACGCTCAGCGGGCTATATACTCGTCGGTGGGGCCG
B18mouse         ----------------TGCTCCGTG-----------CATCGACGTGCTTG
                                  **** * *           * ***  * *   *

B18hum           GCGGTCAGTCTGCGGCAGCGGCAGCAAGACGGTGCAGTGAAGGAGAGTGG
B18mouse         GCAGGCAG--------------AGCAGGCCGG---AAAGAAGCAGGTTGG
                 ** * ***              **** * ***   *  **** **  ***

B18hum           GCGTCTGGCGGGGTCCGCAGTTTCAGCAGAGCCGCTGCAGCCATGGCCCC
B18mouse         GAGTGTGGCGGAGCCCGCAGCTTCAGCAGCTCCGCGGTGACCATGGCCCC
                 * ** ****** * ****** ********  **** *   **********

B18hum           AATCAAGGTGGGAGATGCCATCCCAGCAGTGGAGGTGTTTGAAGGGGAGC
B18mouse         GATCAAGGTGGGAGATGCCATTCCCTCAGTGGAGGTATTTGAAGGGGAAC
                  ******************** **  ********** *********** *

B18hum           CAGGGAACAAGGTGAACCTGGCAGAGCTGTTCAAGGGCAAGAAGGGTGTG
B18mouse         CGGGAAAGAAGGTGAACTTGGCAGAGCTGTTCAAGGGCAAGAAAGGTGTT
                 * ** ** ********* ************************* *****

B18hum           CTGTTTGGAGTTCCTGGGGCCTTCACCCCTGGATGTTCCAAGACACACCT
B18mouse         TTGTTTGGAGTCCCTGGGGCATTTACACCTGGCTGTTCTAAGACCCACCT
                  ********** ******** ** ** ***** ***** ***** *****

B18hum           GCCAGGGTTTGTGGAGCAGGCTGAGGCTCTGAAGGCCAAGGGAGTCCAGG
B18mouse         GCCTGGGTTTGTGGAGCAAGCTGGAGCTCTGAAGGCTAAGGGAGCGCAGG
                 *** ************** ****  *********** *******  ****

B18hum           TGGTGGCCTGTCTGAGTGTTAATGATGCCTTTGTGACTGGCGAGTGGGGC
B18mouse         TGGTGGCCTGTCTGAGCGTTAATGACGTCTTTGTGATTGAAGAGTGGGGT
                 **************** ******** * ******** **  ********

B18hum           CGAGCCCACAAGGCGGAAGGCAAGGTTCGGCTCCTGGCTGATCCCACTGG
B18mouse         CGAGCCCACCAGGCAGAAGGCAAGCTTCGGCTCCTGGCTGACCCCACTGG
                 ********* **** ********* **************** ********

B18hum           GGCCTTTGGGAAGGAGACAGACTTATTACTAGATGATTCGCTGGTGTCCA
B18mouse         AGCCTTTGGGAAGGCGACAGACTTATTATTGGATGATTCTTTGGTGTCTC
                  ************* ************* * ********  *******

B18hum           TCTTTGGGAATCGACGTCTCAAGAGGTTCTCCATGGTGGTACAGGATGGC
B18mouse         TCTTTGGGAATCGTCGGCTGAAAAGGTTCTCCATGGTGATAGACAACGGC
                 ************* ** ** ** *************** ** *  * ***

B18hum           ATAGTGAAGGCCCTGAATGTGGAACCAGATGGCACAGGCCTCACCTGCAG
B18mouse         ATAGTGAAGGCACTGAACGTGGAGCCAGATGGCACAGCCCTCACCTGCAG
                 *********** ***** ***** **************************

B18hum           CCTGGCACCCAATATCATCTCACAGCTCTGAGGCCCTGGGCCAGATTACT
B18mouse         CCTGGCCCCCAACATCCTCTCCCAACTCTGAGGCCCTGG-CCAGATG---
                 ****** ***** *** **** ** ************** ******

B18hum           TCCTCCACCCCTCCCTATCTCACCTGCCCAGCCCTGTGCTGGGGCCCTGC
B18mouse         TCCTCTGACTCTCCC-ATCTCTCCCACCCGGCTCT------AGGCC----
                 *****   * ***** ***** **  *** ** **       ****

B18hum           AATTGGAATGTTGGCCAGATTTCTGCAATAAACACTTGTGGTTTGCGGAA
B18mouse         ----AAAAGGCTCGGTACCTCCTTACTGGGAGC-CACGT-----------
                       ** * * *  *  *   * *    * * *  **
```

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: UNIVERSITE CATHOLIQUE DE LOUVAIN
Halles Universitaires
(B) STREET: Place de l' Universite, 1
(C) CITY: LOUVAIN-LA-NEUVE
(E) COUNTRY: BELGIUM
(F) POSTAL CODE (ZIP): B-1348

(A) NAME: UNIVERSITE DE MONS-HAINAUT
(B) STREET: Place du Parc 20
(C) CITY: MONS
(E) COUNTRY: BELGIUM
(F) POSTAL CODE (ZIP): B-7000

(ii) TITLE OF INVENTION: PEROXISOME-ASSOCIATED PEPTIDE, NUCLEOTIDE SEQUENCE ENCODING SAID PEPTIDE AND THEIR USES IN THE DIAGNOSTIC AND/OR THE TREATMENT OF LUNG INJURIES AND DISEASES, AND OF OXIDATIVE STRESS-RELATED DISORDERS

(iii) NUMBER OF SEQUENCES: 19

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 805 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:193..681

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GCCAGGAGCC GGAGTGGAAG TGGCCGTGGG GCGGGTATGG GACTAGCTGG CGTGTGCGCC 60

CTGAGACGCT CAGCGGGCTA TATACTCGTC GGTGGGGCCG GCGGTCAGTC TGCGGCACCG 120

```
GCAGCAAGAC GGTGCAGTGA AGGAGAGTGG GCGTCTGGCG GGGTCCGCAG TTTCAGCAGA       180

GCCGCTGCAG CC ATG GCC CCA ATC AAG GTG GGA GAT GCC ATC CCA GCA           228
              Met Ala Pro Ile Lys Val Gly Asp Ala Ile Pro Ala
               1               5                  10

GTG GAG GTG TTT GAA GGG GAG CCA GGG AAC AAG GTG AAC CTG GCA GAG         276
Val Glu Val Phe Glu Gly Glu Pro Gly Asn Lys Val Asn Leu Ala Glu
        15                  20                  25

CTG TTC AAG GGC AAG AAG GGT GTG CTG TTT GGA GTT CCT GGG GCC TTC         324
Leu Phe Lys Gly Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe
    30                  35                  40

ACC CCT GGA TGT TCC AAG ACA CAC CTG CCA GGG TTT GTG GAG CAG GCT         372
Thr Pro Gly Cys Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala
45                  50                  55                  60

GAG GCT CTG AAG GCC AAG GGA GTC CAG GTG GTG GCC TGT CTG AGT GTT         420
Glu Ala Leu Lys Ala Lys Gly Val Gln Val Val Ala Cys Leu Ser Val
                65                  70                  75

AAT GAT GCC TTT GTG ACT GGC GAG TGG GGC CGA GCC CAC AAG GCG GAA         468
Asn Asp Ala Phe Val Thr Gly Glu Trp Gly Arg Ala His Lys Ala Glu
            80                  85                  90

GGC AAG GTT CGG CTC CTG GCT GAT CCC ACT GGG GCC TTT GGG AAG GAG         516
Gly Lys Val Arg Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Glu
        95                  100                 105

ACA GAC TTA TTA CTA GAT GAT TCG CTG GTG TCC ATC TTT GGG AAT CGA         564
Thr Asp Leu Leu Leu Asp Asp Ser Leu Val Ser Ile Phe Gly Asn Arg
    110                 115                 120

CGT CTC AAG AGG TTC TCC ATG GTG GTA CAG GAT GGC ATA GTG AAG GCC         612
Arg Leu Lys Arg Phe Ser Met Val Val Gln Asp Gly Ile Val Lys Ala
125                 130                 135                 140

CTG AAT GTG GAA CCA GAT GGC ACA GGC CTC ACC TGC AGC CTG GCA CCC         660
Leu Asn Val Glu Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro
                145                 150                 155

AAT ATC ATC TCA CAG CTC TGA GGCCCTGGGC CAGATTACTT CCTCCACCCC            711
Asn Ile Ile Ser Gln Leu  *
            160

TCCCTATCTC ACCTGCCCAG CCCTGTGCTG GGGCCCTGCA ATTGGAATGT TGGCCAGATT       771

TCTGCAATAA ACACTTGTGG TTTGCGGAAA AAAA                                   805


(2) INFORMATION FOR SEQ ID NO: 2:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 163 amino acids
            (B) TYPE: amino acid
            (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: protein
     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
```

```
Met Ala Pro Ile Lys Val Gly Asp Ala Ile Pro Ala Val Glu Val Phe
  1               5                  10                  15

Glu Gly Glu Pro Gly Asn Lys Val Asn Leu Ala Glu Leu Phe Lys Gly
             20                  25                  30

Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe Thr Pro Gly Cys
         35                  40                  45

Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala Glu Ala Leu Lys
     50                  55                  60

Ala Lys Gly Val Gln Val Val Ala Cys Leu Ser Val Asn Asp Ala Phe
 65                  70                  75                  80

Val Thr Gly Glu Trp Gly Arg Ala His Lys Ala Glu Gly Lys Val Arg
                 85                  90                  95

Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Glu Thr Asp Leu Leu
            100                 105                 110

Leu Asp Asp Ser Leu Val Ser Ile Phe Gly Asn Arg Arg Leu Lys Arg
        115                 120                 125

Phe Ser Met Val Val Gln Asp Gly Ile Val Lys Ala Leu Asn Val Glu
    130                 135                 140

Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro Asn Ile Ile Ser
145                 150                 155                 160

Gln Leu  *
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 780 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus Rattus

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:136..624

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
TGCGTCCTAG GCAGCATAGC CGGATCGGTG CTCCGTGCAT CGGCTACTTG GACGTGCGTG      60

GCAGGCAGAG CAGGCCGGAA AGGAGCAGGT TGGGAGTGTG GTGGGGCCCG CAGCTTCAGC     120
```

```
AGTGCCGCGG TGACTATGGC CCCGATCAAG GTGGGAGACA CCATTCCCTC AGTGGAGGTA    180

TTTGAAGGGG AACCTGGAAA GAAGGTGAAC TTGGCAGAGC TGTTCAAGGA CAAGAAAGGT    240

GTTTTGTTTG GAGTCCCTGG GGCATTTACA CCTGGCTGTT CCAAGACCCA TCTGCCTGGG    300

TTTGTGGAGC AAGCCGGAGC TCYGAAGGCC AAGGGAGCAC AAGTGGTGGC CTGTCTGAGT    360

GTTAATGATG YCTTCGTGAC TGCAGAGTGG GGTCGAGCCC ACCAGGCAGA AGGCAAGGTT    420

CAGCTCCTGG CTGACCCCAC TGGAGCTTTT GGAAAGGAGA CAGATTTACT ACTAGATGAT    480

TCTTTGGTGT CTCTCTTTGG GAATCGTCGG CTAAAAAGGT TCTCCATGGT GATAGACAAG    540

GGCGTAGTAA AGGCACTGAA CGTGGAGCCG GATGGCACAG GCCTCACCTG CAGCCTGGCC    600

CCCAACATCC TCTCACAACT CTGAGGCCCT GACCAGAATG TCCTCTGACT CTCCCATCTC    660

CTCCACCCAG CTCTGGGCCA AAGGCCCAGT ACCTCCTTAC CTGAGGGCCA CTGGAATGGA    720

ACCTTGACAA TATTTCTGCA ATAAACAGTT TAATTTGTGA AAAAAAAAAA AAAAAAAAAA    780
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 162 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus Rattus

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION:17
(D) OTHER INFORMATION:/product= "Glu or Gly"

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION:63
(D) OTHER INFORMATION:/product= "Leu or Pro"

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION:79
(D) OTHER INFORMATION:/product= "Ala or Val"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Ala Pro Ile Lys Val Gly Asp Thr Ile Pro Ser Val Glu Val Phe
1               5                   10                  15
```

```
Xaa Gly Glu Pro Gly Lys Lys Val Asn Leu Ala Glu Leu Phe Lys Asp
            20                  25                  30

Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe Thr Pro Gly Cys
        35                  40                  45

Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala Gly Ala Xaa Lys
    50                  55                  60

Ala Lys Gly Ala Gln Val Val Ala Cys Leu Ser Val Asn Asp Xaa Phe
65                  70                  75                  80

Val Thr Ala Glu Trp Gly Arg Ala His Gln Ala Glu Gly Lys Val Gln
                85                  90                  95

Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Glu Thr Asp Leu Leu
            100                 105                 110

Leu Asp Asp Ser Leu Val Ser Leu Phe Gly Asn Arg Arg Leu Lys Arg
        115                 120                 125

Phe Ser Met Val Ile Asp Lys Gly Val Val Lys Ala Leu Asn Val Glu
    130                 135                 140

Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro Asn Ile Leu Ser
145                 150                 155                 160

Gln Leu
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 675 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mouse

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:99..566

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
TGCTCCGTGC ATCGACGTGC TTGGCAGGCA GAGCAGGCCG GAAAGAAGCA GGTTGGGAGT     60

GTGGCGGAGC CCGCAGCTTC AGCAGCTCCG CGGTGACCAT GGCCCCGATC AAGGTGGGAG    120

ATGCCATTCC CTCAGTGGAG GTATTTGAAG GGGAACCGGG AAAGAAGGTG AACTTGGCAG    180

AGCTGTTCAA GGGCAAGAAA GGTGTTTTGT TTGGAGTCCC TGGGGCATTT ACACCTGGCT    240
```

```
GTTCTAAGAC CCACCTGCCT GGGTTTGTGG AGCAAGCTGG AGCTCTGAAG GCTAAGGGAG      300

CGCAGGTGGT GGCCTGTCTG AGCGTTAATG ACGTCTTTGT GATTGAAGAG TGGGGTCGAG      360

CCCACCAGGC AGAAGGCAAG GTTCGGCTCC TGGCTGACCC CACTGGAGCC TTTGGCAAGG      420

CGACAGACTT ATTATTGGAT GATTCTTTGG TGTCTCTCTT TGGGAATCGT CGGCTGAAAA      480

GGTTCTCCAT GGTGATAGAC AACGGCATAG TGAAGGCACT GAACGTGGAG CCAGATGGCA      540

CAGGCCTCAC CTGCAGCCTG GCCCCCAACA TCCTCTCCCA ACTCTGAGGC CCTGGCCAGA      600

TGTCCTCTGA CTCTCCCATC TCTCCCACCC GGCTCTAGGC CAAAAGGCTC GGTACCTCCT      660

TACTGGGAGC CACGT                                                       675
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 162 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mouse

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Ala Pro Ile Lys Val Gly Asp Ala Ile Pro Ser Val Glu Val Phe
1               5                   10                  15

Glu Gly Glu Pro Gly Lys Lys Val Asn Leu Ala Glu Leu Phe Lys Gly
            20                  25                  30

Lys Lys Gly Val Leu Phe Gly Val Pro Gly Ala Phe Thr Pro Gly Cys
        35                  40                  45

Ser Lys Thr His Leu Pro Gly Phe Val Glu Gln Ala Gly Ala Leu Lys
    50                  55                  60

Ala Lys Gly Ala Gln Val Val Ala Cys Leu Ser Val Asn Asp Val Phe
65                  70                  75                  80

Val Ile Glu Glu Trp Gly Arg Ala His Gln Ala Glu Gly Lys Val Arg
                85                  90                  95

Leu Leu Ala Asp Pro Thr Gly Ala Phe Gly Lys Ala Thr Asp Leu Leu
            100                 105                 110

Leu Asp Asp Ser Leu Val Ser Leu Phe Gly Asn Arg Arg Leu Lys Arg
        115                 120                 125
```

Phe Ser Met Val Ile Asp Asn Gly Ile Val Lys Ala Leu Asn Val Glu
130 135 140

Pro Asp Gly Thr Gly Leu Thr Cys Ser Leu Ala Pro Asn Ile Leu Ser
145 150 155 160

Gln Leu

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 469 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:161..382

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGGTATGGGA CTAGCTGGCG TGTGCGCCCT GAGACGCTCA GCGGGCTATA TACTCGTCGG 60

TGGGGCCGGC GGTCAGTCTG CGGCAGCGGC AGCAAGACGG TGCAGTGAAG GAGAGTGGGC 120

GTCTGGCGGG GTCCGCAGTT TCAGCAGAGC CGCTGCAGCC ATGGCCCCAA TCAAGGTTCG 180

GCTCCTGGCT GATCCCACTG GGGCCTTTGG GAAGGAGACA GACTTATTAC TAGATGATTC 240

GCTGGTGTCC ATCTTTGGGA ATCGACGTCT CAAGAGGTTC TCCATGGTGG TACAGGATGG 300

CATAGTGAAG GCCCTGAATG TGGAACCAGA TGGCACAGGC CTCACCTGCA GCCTGGCACC 360

CAATATCATC TCACAGCTCT GAGGCCCTGG GCCAGATTAC TTCCTCCACC CCTCCCTATC 420

TCACCTGCCC AGCCGTGTGC TGGGGCCCTG CAATTGGAAT GTTGGCCAG 469

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Homo sapiens

(ix) FEATURE:
    (A) NAME/KEY: CDS
    (B) LOCATION:161..514

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
GGGTATGGGA CTAGCTGGCG TGTGCGCCCT GAGACGCTCA GCGGGCTATA TACTCGTCGG    60

TGGGGCCGGC GGTCAGTCTG CGGCAGCGGC AGCAAGACGG TGCAGTGAAG GAGAGTGGGC   120

GTCTGGCGGG GTCCGCAGTT TCAGCAGAGC CGCTGCAGCC ATGGCCCCAA TCAAGACACA   180

CCTGCCAGGG TTTGTGGAGC AGGCTGAGGC TCTGAAGGCC AAGGGAGTCC AGGTGGTGGC   240

CTGTCTGAGT GTTAATGATG CCTTTGTGAC TGGCGAGTGG GGCCGAGCCC ACAAGGCGGA   300

AGGCAAGGTT CGGCTCCTGG CTGATCCCAC TGGGGCCTTT GGGAAGGAGA CAGACTTATT   360

ACTAGATGAT TCGCTGGTGT CCATCTTTGG GAATCGACGT CTCAAGAGGT TCTCCATGGT   420

GGTACAGGAT GGCATAGTGA AGGCCCTGAA TGTGGAACCA GATGGCACAG GCCTCACCTG   480

CAGCCTGGCA CCCAATATCA TCTCACAGCT CTGAGGCCCT GGGCCAGATT ACTTCCTCCA   540

CCCCTCCCTA TCTCACCTGC CCAGCCCTGT GCTGGGGCCC TGCAATTGGA ATGTTGGCCA   600

G                                                                   601
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 604 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Homo sapiens

(ix) FEATURE:
    (A) NAME/KEY: CDS
    (B) LOCATION:161..517

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
GGGTATGGGA CTAGCTGGCG TGTGCGCCCT GAGACGCTCA GCGGGCTATA TACTCGTCGG    60
```

TGGGGCCGGC GGTCAGTCTG CGGCAGCGGC AGCAAGACGG TGCAGTGAAG GAGAGTGGGC 120

GTCTGGCGGG GTCCGCAGTT TCAGCAGAGC CGCTGCAGCC ATGGCCCCAA TCAAGGTGGG 180

AGATGCCATC CCAGCAGTGG AGGTGTTTGA AGGGGAGCCA GGGAACAAGG TGAACCTGGC 240

AGAGCTGTTC AAGGGCAAGA AGGGTGTGCT GTTTGGAGTT CCTGGGGCCT TCACCCCTGG 300

ATGTTCCAAG GTTCGGCTCC TGGCTGATCC CACTGGGGCC TTTGGGAAGG AGACAGACTT 360

ATTACTAGAT GATTCGCTGG TGTCCATCTT TGGGAATCGA CGTCTCAAGA GGTTCTCCAT 420

GGTGGTACAG GATGGCATAG TGAAGGCCCT GAATGTGGAA CCAGATGGCA CAGGCCTCAC 480

CTGCAGCCTG GCACCCAATA TCATCTCACA GCTCTGAGGC CCTGGGCCAG ATTACTTCCT 540

CCACCCCTCC CTATCTCACC TGCCCAGCCC TGTGCTGGGG CCCTGCAATT GGAATGTTGG 600

CCAG 604

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2710 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:2516..2710

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:2074..2135

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:1932..1970

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:1728..1859

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION:802..936

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
TCTGTCCCTT AGCGCCCCCG CGGGGGCTTA CCCCATCCCA CTCCATGACC TCCCCTCCCC     60
CCATGGCGAA TTCCCACCTT TCTGTCTTTC ACTCACTTCC TGGAACCGTC CCCAGGGCCT    120
TGGACCTTCC CCCTTCTCCT CCCAAACCTT GTGAGACCCC ATTCCCTTTC TACTTCATCC    180
TGCTCTCAAC TTTTGGGCTC CTCAGAGGCC CTCACCCCTG ACTCTCTCTC CCTACCCACT    240
CTGGTCCCAT GAAGCCCTCA AGTACTCTGG GGATGGATCC TTCCCCCTTC AAAAGATTCC    300
TTCTTTTGTT CTACACCTCC TGGGTGTAGG GGCCTGGACA CCCTCCCCCA ACGTTCCACC    360
TGCCGCTGCC CTTCCTCTTC CTCCTCCTGA GGGTGGGACC CTCAGACCTG GCCAAGATCC    420
TCTCCCTCCA TGTTGTCAGG GACTCCTCCT CACCCCCAAA TACAGCCCTC TAGCCCCTGT    480
CCATTTTATT CCACTCCTTT CCTGTAACCT AGACAGCATG TTATGCAACC CTTTGCGACA    540
CATGGGGAAA CCTTCCCTCC CTTCCTCTGT TGTCACCAAT GGCCCCTTAA GAGGAGCAGG    600
GCCACCTTGA AACTTGGAGG ATATGGGGTA ACCCAGTGGG AGCGGGCAGG GAGGGCCCTT    660
GGAAACTGAC AGGGCTGGAG TATCCTGCTG GGTTTCAGCC CCGGTTCCTG CAGGCACAGC    720
TGCCAGGCTC TCTGTTCACC TTCCTGCCTC TGGTTTGCCC CGGCTCCCTC ACCCCCCTTA    780
CCCTGGAGTC CTTCCTTCTA GGTGGGAGAT GCCATCCCAG CAGTGGAGGT GTTTGAAGGG    840
GAGCCAGGGA ACAAGGTGAA CCTGGCAGAG CTGTTCAAGG GCAAGAAGGG TGTGCTGTTT    900
GGAGTTCCTG GGGCCTTCAC CCCTGGATGT TCCAAGGTGA GGCCCTTCCC CTTCTGAAGA    960
TCAGGACCTG GGGATCTTTT GTGTTGCTCT TAAGTCCTCC ACATAGTCCT GATAGGACTC   1020
CTAAAAAGCA TTTCAGTGCC ATCACAAAAC AAGTAGAGCT GGGTAGAGCT GGGCGCGGTG   1080
GCTCACGCCT GTAATCCCAG CACTTTGGGA GGCCAAGGCG GGTGGATCAC GAGGTCAGGA   1140
GTCCAAAACC AGCCTGGCCA AGATGGTGAA ACCCTGTCTC TACTAAAAAT GCAAAAAAAT   1200
CAGCCGGATA TGGTGGCGGG CGCCTGTAAT CCCAGGTATT GGGGAGGCTG AGGCAGAGAA   1260
TTGCTTGAAC CCAGGAGGCG TAGGTTGCAG TGAGTGGAGA TCGTGCCTCT GCAGTCCAGC   1320
CTGGGTGAAA GAGCGAGACT CCGTCTCAAA ATGAAAAAAA AAAAAGAAAA CAAGTAGAGA   1380
CTGCAAAAAG GGAACAGTAC CGGGAATGTT GGAGAAAAAC ATACTACAAT TAAATCCAAC   1440
ACCCCTGTTG GTCCTGCTAA ATGACAGGCA CTGTGGAAGG TGCTTGGGAC TCAGATAAAT   1500
AAGACAAAGA TCTGCCCATG GAAAGTTCAC GTCTGGACCA TAAGGCATTA GGTTTCATTC   1560
TGAGCTTCCT AGTGGCCAAG GCAAAAAGGA AATAGAATGG TTTAGACAGC TCTCATTGTC   1620
TGATCAAAGG TGTTGAGGCA GAGCACTGAG GAGGGCCTGG AGATAAAGGG TGGGCTGGGG   1680
GTCAGATGCA GTTATCCCTT TGCCGACCCT TTGTTCCCCT TCCTCAGACA CACCTGCCAG   1740
GGTTTGTGGA GCAGGCTGAG GCTCTGAAGG CCAAGGGAGT CCAGGTGGTG GCCTGTCTGA   1800
GTGTTAATGA TGCCTTTGTG ACTGGCGAGT GGGGCCGAGC CCACAAGGCG GAAGGCAAGG   1860
```

```
TGAGGTGAGG GGCCTGCAGG GAGTCAGGAC CAGGTAGGAT ATTCTTCTTG TGACCTCTAC    1920

TTTCTCTGCA GGTTCGGCTC CTGGCTGATC CCACTGGGGC CTTTGGGAAG GTGAGTGTTC    1980

CCCTGACCGC CACAGGGACA TGGCGGTGCG GGGAGCAGTG GGGGCCCTTG GCCTCTTCAA    2040

GGATTTCTGA CACTTTTCTC TGTCTCTTCT TAGGAGACAG ACTTATTACT AGATGATTCG    2100

CTGGTGTCCA TCTTTGGGAA TCGACGTCTC AAGAGGTAAA AGTGGAGAGT CCTCTGTGGA    2160

GAAAGTCCTC TGTGGGAGAG AGTCCTCTGT GGGAGAGAGT CCTCTGTGGA GAGGGTCCTC    2220

TGTGGGAAGA GTCGTCTGTG GGGGAGATGT GTGGGAGAGA GTCCTGTGTG GGGAGAGTCT    2280

TCTGTAGGGG AGAGTCCTCT GGGGAGAGAG TCCTGTGTGG GGGAGAGTCC TCTGTGGGGA    2340

GAGTCCTCTG TGTGGAGAGA GTCCTGTGTG GTGGTGAGTC CTCTGTGGGG GAGAGTCCTC    2400

TGTGGGGGGA GTCCTCTCTG GAGTTCTCTT GGGCCCCTGG CTGTTCACTG CCTGTCTCCA    2460

TGCCCAGCCT CCAAGCCCAG GCTGATGCAG CTGGCTGGGC CCCTCTTTCC GGCAGGTTCT    2520

CCATGGTGGT ACAGGATGGC ATAGTGAAGG CCCTGAATGT GGAACCAGAT GGCACAGGCC    2580

TCACCTGCAG CCTGGCACCC AATATCATCT CACAGCTCTG AGGCCCTGGG CCAGATTACT    2640

TCCTCCACCC CTCCCTATCT CACCTGCCCA GCCCTGTGCT GGGGCCCTGC AATTGGAATG    2700

TTGGCCAGAT                                                           2710
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
GCCATCCCAG CAGTGGAGGT GTTTG                                          25
```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TTGAACAGCT CTGCCAGGTT CACC 24

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TGGAGGTGTT TGAAGGGGAG CCAG 24

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CAGGTTCACC TTGTTCCCTG GCTC 24

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GGGTATGGGA CTAGCTGGCG 20

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CTGGCCAACA TTCCAATTGC AG 22

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATGTTATGCA ACCCTTTGCG ACAC 24

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GTGTTTGAAG GGGAGCCAGG GAAC 24

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGAGACAGGG TTTCACCATC TTGG 24

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : **C12N 15/12, C07K 14/47**	**A3**	**(11) International Publication Number:** **WO 99/09054** **(43) International Publication Date:** 25 February 1999 (25.02.99)

(21) International Application Number: PCT/BE98/00124

(22) International Filing Date: 20 August 1998 (20.08.98)

(30) Priority Data:
9700692 20 August 1997 (20.08.97) BE

(71) Applicants *(for all designated States except US)*: UNIVERSITE CATHOLIQUE DE LOUVAIN [BE/BE]; Halles Universitaires, Place de l'Université 1, B–1348 Louvain–la–Neuve (BE). UNIVERSITE DE MONS–HAINAUT [BE/BE]; Place du Parc 20, B–7000 Mons (BE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KNOOPS, Bernard [BE/BE]; Rue Chapelle Notre–Dane 3/1, B–1341 Ceroux–Mousty (BE). HERMANS, Cedric [BE/BE]; Avenue des Glycines 42, B–1030 Brussels (BE). BERNARD, Alfred [BE/BE]; Avenue de la Chapelle 6, B–1200 Brussels (BE). WATTIEZ, Ruddy [BE/BE]; Chemin du Sauvelon 17, B–7022 Hyon (BE). FALMAGNE, Paul [BE/BE]; Rue du Point du Jour 8, B–7022 Mesvin (BE).

(74) Agents: VAN MALDEREN, Eric et al.; Office Van Malderen, Place Reine Fabiola 6/1, B–1083 Brussels (BE).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(88) Date of publication of the international search report: 15 April 1999 (15.04.99)

(54) Title: PEROXISOME–ASSOCIATED POLYPEPTIDE, NUCLEOTIDE SEQUENCE ENCODING SAID POLYPEPTIDE AND THEIR USES IN THE DIAGNOSIS AND/OR THE TREATMENT OF LUNG INJURIES AND DISEASES, AND OF OXIDATIVE STRESS–RELATED DISORDERS

(57) Abstract

The present invention is related to an isolated and purified polypeptide which amino acid sequence presents more than 70 % with the sequence SEQ ID NO 1. The present invention is also related to the nucleotide sequence encoding said amino acid sequence, the inhibitor directed against said sequences and their use in the diagnosis, treatment and/or prevention of lung injuries or diseases and oxidative stress-related disorders.

```
CLUSTAL V alignment of human and rat B18 amino acid sequences (Identity:
90%, Homology: 97.5%):

B18hum     MAPIKVGDAIPAVEVFEGEPGNKVNLAELFKGKKGVLFGVPGAFTPGCSK = SEQIDNO1
B18rat     MAPIKVGDTIPSVEVFEGEPGKKVNLAELFKDKKGVLFGVPGAFTPGCSK
           ********.**.*********.*********.******************

B18hum     THLPGFVEQAEALKAKGVQVVACLSVNDAFVTGEWGRAHKAEGKVRLLAD
B18rat     THLPGFVEQAGALKAKGAQVVACLSVNDVFVTAEWGRAHQAEGKVQLLAD
           ********** ****** ********** ***.******.*****.****

B18hum     PTGAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQDGIVKALNVEPDGTGL
B18rat     PTGAFGKETDLLLDDSLVSLFGNRRLKRFSMVIDKGVVKALNVEPDGTGL
           *******************.************.. *.*************

B18hum     TCSLAPNIISQL
B18rat     TCSLAPNILSQL
           ********.***


CLUSTAL V alignment of human and mouse B18 amino acid sequences (Identity:
91%, Homology: 96%):

B18hum     MAPIKVGDAIPAVEVFEGEPGNKVNLAELFKGKKGVLFGVPGAFTPGCSK
B18mouse   MAPIKVGDAIPSVEVFEGEPGKKVNLAELFKGKKGVLFGVPGAFTPGCSK
           ***********.*********.****************************

B18hum     THLPGFVEQAEALKAKGVQVVACLSVNDAFVTGEWGRAHKAEGKVRLLAD
B18mouse   THLPGFVEQAGALKAKGAQVVACLSVNDVFVIEEWGRAHQAEGKVRLLAD
           ********** ****** ********** **  ******.**********

B18hum     PTGAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQDGIVKALNVEPDGTGL
B18mouse   PTGAFGKATDLLLDDSLVSLFGNRRLKRFSMVIDNGIVKALNVEPDGTGL
           ******* ***********.************...***************

B18hum     TCSLAPNIISQL
B18mouse   TCSLAPNILSQL
           ********.***


CLUSTAL V alignment of human and rat cDNA sequences (identity: 612/780,
78.5%):

B18hum      GCCAGGAGGCGGAGTGGAAGTGGCCGTGGGGCGGGTATGGGACTAGCTGG
B18rat      --------------TG--------CGTC--------------CTAGGCAG
                          **        ***               ****   *

B18hum      CGTGTGCGCCCTGAGACGCTCAGCGGGCTATATACTCGTCGGTGGGGCCG
B18rat      CATA---GCC---GGA---TCGGTGCTCCGTGCATCGGCTACTTGGAC--
            * *    ***   **    ** * *  *  *  *  *    * ** *

B18hum      GCGGTCAGTCTGCGGCAGCGGCAGCAAGACGGTGCAGTGAAGGAGAGTGG
B18rat      -------GTGCGTGGCAGGCAGAGCAGGCCGG---AAAGGAGCAGGTTGG
                   **  * *****    **** * ***   *  * ** **  ***
```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

INTERNATIONAL SEARCH REPORT

Inter. .nal Application No
PCT/BE 98/00124

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL Accssion number N42215, 27 January 1996 HILLIER L. ET AL.: "N42215 I64154 hypothetical protein HI0572 - Haemophilus influenzae" XP002089366 see abstract ---	8,9
P,X	DATABASE EMBL Accession number AA639364, 1 November 1997 STRAUSBERG R.: "nq87d08.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1159311 similar to TR:G558080 G558080 PUTATIVE PEROXISOMAL PROTEIN." XP002089367 see abstract --- -/--	8,9

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search 7 January 1999	Date of mailing of the international search report 12. 02. 99
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Chakravarty, A

2

INTERNATIONAL SEARCH REPORT

Inter. .nal Application No
PCT/BE 98/00124

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DATABASE EMBL Accession number P73728, 1 February 1997 TABATA S.: "MEMBRANE PROTEIN from SYNECHOCYSTIS SP." XP002089368 see abstract -----	1-4

2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/BE 98/00124

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 10, 11-15 (part)
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claims Nos.: 10, 11-15 (part)

Claim 10 is directed to inhibitors of the amino acid / nucleotide sequences of claims 1-8. However, no such compounds are defined in the application. No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the required result.
The same objection applies, mutatis mutandis, to claims 11-15 in as much as these claims refer to the subject-matter of claim 10.